

Dynamic System Optimum Analysis of Multi-Region Macroscopic Fundamental Diagram Systems With State-Dependent Time-Varying Delays

Renxin Zhong^{ID}, Jianhui Xiong, Yunping Huang, Agachai Sumalee, Andy H. F. Chow^{ID}, and Tianlu Pan

Abstract—This paper investigates the dynamic system optimum (DSO) problem with simultaneous route and departure time assignments for a general traffic network partitioned into multiple regions. Regional traffic congestion is modeled with a well-defined macroscopic fundamental diagram (MFD) mapping the trip completion rate to the vehicular accumulation. To overcome the limitation of inconsistent flow propagation between region boundaries and the corresponding travel time, the state-dependent regional travel time function is explicitly incorporated in the flow propagation of the conventional MFD dynamics. From a systems perspective, the traffic dynamics within a region can be regarded as a dynamic system with an endogenous time-varying delay depending on the system state. Equilibrium condition for the DSO problem is analytically derived through the lens of Pontryagin minimum principle and is compared against the static SO counterpart. The structure of path specific marginal cost is analyzed regarding the path travel cost and early-late penalty function. In contrast to existing analytical methods, the proposed method is applicable for general MFD systems without linearization of the MFD dynamics. Neither approximation of the equilibrium solution nor constant regional delay assumption is required. Numerical examples are conducted to illustrate the characteristics of DSO traffic equilibrium and the corresponding marginal cost together with other dynamic external costs.

Index Terms—Dynamic system optimum, macroscopic fundamental diagram, state-dependent time-varying delay, simultaneous route assignment and departure time assignment, saturated state and input constraints.

Manuscript received April 2, 2019; revised October 30, 2019 and March 22, 2020; accepted May 8, 2020. Date of publication May 27, 2020; date of current version August 28, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1600500, in part by the Research Grants Council of Hong Kong SAR under Project 15210117E and Project 15211518E, and in part by the CCF-DiDi Big-Data Joint Lab. The Associate Editor for this article was J. Sanchez-Medina. (*Corresponding author: Tianlu Pan.*)

Renxin Zhong and Jianhui Xiong are with the Guangdong Key Laboratory of Intelligent Transportation Systems, School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 510006, China.

Yunping Huang is with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong.

Agachai Sumalee is with the School of Integrated Innovation, Chulalongkorn University, Bangkok 10330, Thailand.

Andy H. F. Chow is with the Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong.

Tianlu Pan is with the Guangdong Key Laboratory of Urban Informatics, School of Architecture and Urban Planning, Research Institute for Smart Cities, Shenzhen University, Shenzhen 518061, China (e-mail: glorious9009@gmail.com).

Digital Object Identifier 10.1109/TITS.2020.2994347

I. INTRODUCTION

THE system optimal (SO) problem that minimizes the total travel time spent by all travelers in the network is regarded as a benchmark for assessing the overall performance of a traffic network. Dynamic extension of the SO problem, i.e., DSO problem, analyzes traffic equilibrium under which the user surplus is maximized [1]–[4]. The DSO equilibrium provides a bound on the best performance of a dynamic traffic network, which makes it as a benchmark for evaluating various congestion management measures, e.g., dynamic road pricing and tradable network permit [2], [5], network traffic access control [3], [6], [7], and dynamic road capacity allocation [8], [9]. DSO problems are formulated according to the combination of route choice (or assignment) and departure time choice (or assignment), i.e., the departure time assignment and the morning commute problems (e.g., [10]), the route assignment problems (e.g., [11]–[16]), and the simultaneous route and departure time assignments (e.g., [3], [5], [17]–[19]).

The DSO equilibrium highly depends on the underlying dynamic network loading models that capture traffic flow propagation over the network. Depending on how and to which level of capturing the traffic congestion, typical link-level dynamic network loading models can be grouped into two categories: 1) the vertical queue models which assume vehicles queue vertically, e.g. deterministic bottleneck model or the point queue models (e.g., [3], [20]), and the whole link models (e.g., [3], [11], [12], [21]); 2) the physical queue models that consider the queue length and congestion wave, e.g., the kinematic wave model and its discretization the cell transmission model (e.g., [13], [22]–[24]), and the link transmission model [4]. Physical queue models would better capture the processes of traffic congestion onset and dissipation than vertical queue models. However, they would be more computationally intensive. Nevertheless, these models would suffer from the scalability problem for large network applications, i.e., a large spatial dimension.

The macroscopic fundamental diagram (MFD) gives rise to a promising solution to the challenge of spatial dimensionality. The MFD describes the aggregate traffic network dynamics under stationary and homogenous traffic assumption by mapping the network flow accumulation to the trip completion rate. The traffic network within each urban region can be regarded

as a macro-scale queueing system with a time-dependent inflow [25]. Daganzo [26] and Daganzo and Geroliminis [27] analytically formulated the MFD dynamics based on the variational theory. Geroliminis and Daganzo [28] presented an empirical calibration of the MFD for the city of Yokohama in Japan. The asymmetric OD, the route choice behavior, and the randomness of turning movements are regarded as the major causes of scatters in MFDs [29]–[31]. The mechanism of network configuration and route choice behaviors on the throughput was further investigated in [32]. Zhong *et al.* [33] investigated the effect of time-varying demand on the behavior of the MFD dynamics and proposed some appropriate boundary conditions to ensure the proper posedness of MFD dynamics. A fundamental issue of controllability via perimeter control was investigated in Zhong *et al.* [33] wherein a set of sufficient conditions that guarantee the controllability are derived for general multi-region MFD systems.

However, travelers' route choice and departure time choice, two important factors regarding traffic control, are missing in these studies regardless of their importance in the existence of the MFD. Yildirimoglu and Geroliminis [34] integrated the MFD dynamics with aggregated route choice dynamics to develop a region-based route choice model using hierarchical control framework. Approximate dynamic user equilibrium (DUE) and DSO conditions for route choice behaviors under the MFD framework were further pursued by Yildirimoglu *et al.* [35]. Amirgholy and Gao [25] applied the MFD to describe the network traffic dynamics in the morning commute problem (single MFD region without route choice behavior) to enable departure time choice. Dynamic tolling strategy for network users inside the region was devised to minimize the generalized cost of the system via keeping the outflow maximized over the peak. On the other hand, by assuming a prescribed departure curve, Ampountolas *et al.* [36] adopted the C-Logit model to describe the route choice behavior to simulate drivers' adaptiveness to traffic conditions with the presence of a robust perimeter and boundary flow controller. Yildirimoglu *et al.* [37] proposed a hierarchical traffic management scheme under the MFD framework by integrating the regional route guidance schemes in the upper-level via optimizing regional split ratios and the path assignment mechanism in the lower-level via recommending subregional paths for vehicles to follow. Aghamohammadi and Laval [38] stated that further research efforts should be dedicated to addressing the departure time choice modeling and capacity constraint under the MFD framework. The DUE problem considering simultaneous route choice and departure time choice under the MFD framework was investigated in our recent research [39]. The DSO with simultaneous route and departure time assignments for a general traffic network modeled by multi-region MFD systems is still unexplored.

It is postulated in the above-mentioned literature that the MFD dynamics is described by a set of ordinary differential equations wherein vehicles can immediately traverse the region and complete its trip, i.e., the travel time from the region to its border is negligible [40], [41]. Since vehicles cannot move immediately from one of the regional borders to another but would exit the network after some delays equal to the travel

times required to traverse the network, the regional travel time as a function of the network accumulation should be explicitly integrated into the MFD system dynamics. The travel time is regarded as a time delay in the control input by [40], [42], [43] wherein the control input delay is assumed to be constant in the MFD dynamics for perimeter control design to simplify the model description. Haddad and Zheng [41] incorporated the constant time delay in the state dynamics to better capture the traffic flow propagation using the MFD dynamics to improve the adaptive perimeter control strategies. However, since the network accumulation is a dynamic process, state-dependent travel time is thus time-varying and endogenous due to the dynamic traffic conditions such as congestion onset and dissolve and the fast time-varying nature of travel demand. How to handle such kind of endogenous time-varying delay subject to state and input constraints is still missing in transportation literature and is marked as a future research topic by Haddad and Zheng [41].¹ Therefore, we extend the conventional MFD dynamics to explicitly incorporate the endogenous time-varying delays in the flow propagation. Nevertheless, conventional local linearization technique that linearizes the system dynamics at a given equilibrium point is often adopted in the MFD literature, see e.g., [40]–[43]. As discussed in [33], [46], the local linearization technique can only look into the local properties around the reference point to which the linearization is performed. To avoid the hurdle caused by the local linearization and constant delay assumption, we stick to the nonlinear MFD dynamics with state-dependent delays and saturated input constraints.

To tackle the aforementioned challenges, this paper proposes a formulation for the unexplored DSO problem considering simultaneous route and departure time assignments under the MFD framework. It is assumed that there is a central traffic manager to coordinate all commuters' departure times and routes to minimize the total system cost to achieve a DSO equilibrium. The inflow capacity constraints are captured by certain inequality constraints. An optimal control formulation is developed to derive the necessary condition for the DSO equilibrium analytically under the umbrella of the Pontryagin minimum principle. The structures of the dynamic marginal travel cost and other external cost are analyzed for dynamic tolling design. Technically, the paper contributes to the state-of-the-art of the MFD framework: 1) by extending the MFD dynamics to incorporate the travel time function explicitly for proper flow propagation; 2) introducing an analytical capacity constraint to ensure that the inflow rate to a region cannot exceed its maximal allowance; 3) abandoning the local linearization technique often adopted in the MFD literature.

Paper organization: Section II introduces the MFD model with state-dependent time-varying delay for network loading. To model simultaneous route and departure time assignments, Section III formulates the DSO problem via optimal control with inequality constraints. Equilibrium condition is derived through the lens of Pontryagin minimum principle. Section IV conducts numerical examples to illustrate the properties of

¹Note that another way to model the regional travel delay in the MFD dynamics is to use the trip-based MFD models, see e.g., [44], [45].

TABLE I
NOMENCLATURE

\mathcal{W}	The set of all OD pairs
\mathcal{P}	The set of all paths
w	OD pair w . Generally refers to any OD pairs in \mathcal{W}
p	Path p . Generally refers to any path in \mathcal{P}
R	Region R . Generally refers to any region in \mathcal{R}
R_j	'Physical' region R_j , the j^{th} region of all regions
R_i^p	'Virtual' region R_i^p , the i^{th} region along path p , $i \in [1, m(p)]$
\mathcal{R}	The set of all 'physical' regions, and $R, R_j \in \mathcal{R}$
\mathcal{R}^p	The set of all regions along path p , $\mathcal{R}^p \subseteq \mathcal{R}$, and $R_i^p \in \mathcal{R}^p$
δ_{R_j, R_i^p}	The Kronecker Delta function. In this paper, we use this function to describe the relation between the elements of \mathcal{R} and \mathcal{R}^p : If the i^{th} region along path p , i.e., 'virtual' region $R_i^p \in \mathcal{R}^p$, is the 'physical' region $R_j \in \mathcal{R}$, then $\delta_{R_j, R_i^p} = 1$. Otherwise, $\delta_{R_j, R_i^p} = 0$, i.e., $\delta_{R_j, R_i^p} = \begin{cases} 1, & \text{if } R_i^p = R_j \\ 0, & \text{otherwise} \end{cases}$
\mathcal{G}_R	Network output (or exit) function of region R , also known as the MFD
\mathcal{P}_w	The set of all paths connecting OD pair w , $\mathcal{P}_w \subseteq \mathcal{P}$
$m(p)$	The number of regions along path p , i.e., $m(p) = \mathcal{R}^p $
Q_w	Travel demand of OD pair w
ϕ_w	Minimum travel cost of OD pair w
$q^p(t)$	Departure rate of path p at time t
$G_R^p(t)$	Path-based flow (rate) exiting region R along path p at time t . $G_R^p(t)$ denotes the flow exiting region R_i^p along path p at time t
$n_R^p(t)$	Traffic accumulation contributed by path p of region R at time t . $n_R^p(t)$ denotes the traffic accumulation contributed by path p of region R_i^p at time t
$n_R(t)$	Traffic accumulation state of region R at time t . $n_R(t) = \sum_{p \in \mathcal{P}} \sum_{i=1}^{m(p)} n_R^p(t) \delta_{R, R_i^p}$
$h_R(\cdot)$	Travel time (or delay) of region R with $h_R(n_R) _{n_R \rightarrow 0}$ denoting the free-flow travel time
$\sigma_R(\cdot)$	Experienced or inverse travel time of region R
$\tau_R^p(t)$	Exit time from region R of vehicles departing from the origin at time t along path p . $\tau_R^p(t)$ denotes the exit time from the region R_i^p (also the entry time to the region R_{i+1}^p) along path p
$h^p(\cdot)$	Journey time (or path delay operator) of path p
$\Psi^p(\cdot)$	Effective travel cost of path p including schedule delay penalty

the DSO equilibrium and the dynamic marginal cost. Finally, Section V concludes the paper. For convenience of readers, we summarize the nomenclature used in this paper in Table I. Companion materials such as the solution algorithms are presented in the online appendix.²

II. MULTI-REGION MACROSCOPIC FUNDAMENTAL DIAGRAM SYSTEMS WITH STATE DEPENDENT TIME-VARYING DELAYS

An urban traffic network is partitioned into M regions, labeled as $R_1, R_2, \dots, R_j, \dots, R_M$, on condition that the traffic is homogeneously distributed within each region such that each region admits a well-defined MFD. For a region R , the production-MFD, i.e., $P_R(n_R)$, is a concave function of the network accumulation $n_R(t)$ (veh) (the total number of vehicles of region R at time t). Assuming average trip length in region R is L_R , the average travel time function of region R is $h_R(n_R) = \frac{L_R}{v_R(n_R)}$ with v_R is the network average space-mean velocity given by $v_R(n_R(t)) = P_R(n_R(t))/n_R(t)$. The network output (or exit) function, also known as the MFD, can be deduced as $\mathcal{G}_R(n_R) = \frac{P_R(n_R)}{L_R}$. By assumption $\mathcal{G}_R(n_R)$ (veh/s) is a concave function of $n_R(t)$. Therefore, the average network travel time function of region R at time t can be evaluated as a function of $n_R(t)$ at time t as

$$h_R(n_R(t)) = \frac{L_R}{v_R(n_R(t))} = \frac{n_R(t)}{\frac{P_R(n_R(t))}{L_R}} = \frac{n_R(t)}{\mathcal{G}_R(n_R(t))} \quad (1)$$

²The online appendix is available at https://www.dropbox.com/s/3wj0ld4n3nk3yf/Companion%20materials%20of%20DSO_IEEE%20ITSC.pdf?dl=0.

When $n_R(t) = 0$, the network is under the free flow condition. Thus the velocity $v_R(0)$ is the free-flow speed, and travel time $h_R(0)$ is the free-flow time. By (1), the free flow travel time of region R can be obtained by applying L'Hôpital's rule

$$h_R(n_R)|_{n_R \rightarrow 0} = \lim_{n_R \rightarrow 0} \frac{n_R}{\mathcal{G}_R(n_R)} = \lim_{n_R \rightarrow 0} \frac{\dot{n}_R}{\dot{\mathcal{G}}_R(n_R)} = \frac{1}{\dot{\mathcal{G}}_R(0)}$$

where we have omitted t for simplicity.

For a single-region MFD system, its dynamics evolves according to the flow conservation:

$$\frac{dn_R(t)}{dt} = q_R(t) - G_R(t)$$

where $q_R(t)$ denotes the inflow rate to region R , and $G_R(t)$ denotes the outflow (or exit) rate from region R at time t . Or equivalently,

$$n_R(t) = n_R(0) + \int_0^t (q_R(s) - G_R(s)) ds \quad (2)$$

where $n_R(0)$ denotes the initial network accumulation of region R . Without loss of generality, we assume $n_R(0) = 0$. This simple model postulates that the inflow is transmitted to outflow boundary at an infinite speed or with a negligible travel time. However, vehicles cannot immediately traverse a region nor move from one to another. The travel time as a function of the network accumulation, should be inherently integrated in the MFD system dynamics.

To consider the travel time in flow propagation, we define the "experienced or inverse travel time" function $\sigma_R(t)$ as the average travel time of vehicles exiting the region at time t .

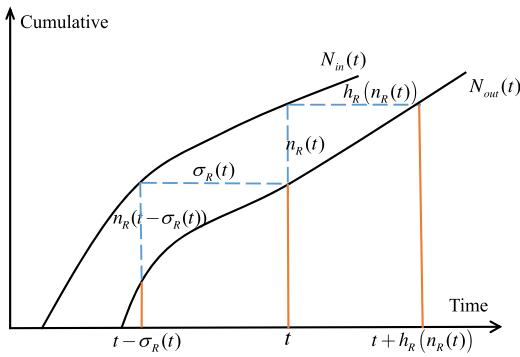


Fig. 1. Relationship between time index and travel time functions.

That is vehicles entering region R at time $t - \sigma_R(t)$ would exit from the region at time t . Referring to Fig. 1, in terms of flow conservation, we write $N_{out}(t) = N_{in}(t - \sigma_R(t))$, where $N_{in}(t)$ and $N_{out}(t)$ are cumulative inflow to region R and cumulative outflow from region R up to time t , respectively. By incorporating this experienced travel time definition into the flow conservation, we have the following relationship:

$$N_{in}(t - \sigma_R(t)) = \int_0^{t - \sigma_R(t)} q_R(s) ds = N_{out}(t) = \int_0^t G_R(s) ds \quad (3)$$

Differentiating (3) yields:

$$G_R(t) = (1 - \dot{\sigma}_R(t)) q_R(t - \sigma_R(t)) \quad (4)$$

However, the travel time function is a more intuitive measure in the literature. Following [44], [47], the flow conservation using the travel time function $h_R(n_R(t))$ for vehicles entering region R at time t is defined as

$$N_{out}(t + h_R(n_R(t))) = N_{in}(t). \quad (5)$$

That is,

$$\int_0^t q_R(s) ds = \int_0^{t + h_R(n_R(t))} G_R(s) ds \quad (6)$$

Differentiating (6) yields

$$(1 + \dot{h}_R(n_R(t))) G_R(t + h_R(n_R(t))) = q_R(t)$$

Since both $G_R(\cdot)$ and $q_R(\cdot)$ are nonnegative, $1 + \dot{h}_R(n_R(t)) > 0$ so as to make this nonnegative be fulfilled. In the dynamic traffic assignment (DTA) literature, this condition is known as First-In-First-Out (FIFO) principle. If $1 + \dot{h}_R(n_R(t)) > 0$, we have

$$G_R(t + h_R(n_R(t))) = \frac{q_R(t)}{1 + \dot{h}_R(n_R(t))}. \quad (7)$$

The differences between the experienced travel time function and the travel time function, as well as the calculation routines were discussed in [48]. Although (4) is obtained from experienced travel time and (7) is obtained from travel time, the two relationship are mathematically equivalent. To see this, note that vehicles entering the region at time $t - \sigma_R(t)$

输出率和输入率之间的关系

exit from the region R at time t , by the definition of travel time we have

$$\sigma_R(t) = h_R(t - \sigma_R(t))$$

Differentiating the above equation yields:

$$\begin{aligned} \dot{\sigma}_R(t) &= \frac{dh_R(t - \sigma_R(t))}{dt} = \frac{dh_R(t - \sigma_R(t))}{d(t - \sigma_R(t))} \frac{d(t - \sigma_R(t))}{dt} \\ &= \frac{dh_R(t - \sigma_R(t))}{d(t - \sigma_R(t))} (1 - \dot{\sigma}_R(t)) \end{aligned}$$

Then we have

$$\dot{\sigma}_R(t) = \frac{dh_R(t - \sigma_R(t))/d(t - \sigma_R(t))}{1 + dh_R(t - \sigma_R(t))/d(t - \sigma_R(t))}$$

Substituting this into (4) yields:

$$G_R(t) = (1 - \dot{\sigma}_R(t)) q_R(t - \sigma_R(t)) = \frac{q_R(t - \sigma_R(t))}{1 + \frac{dh_R(t - \sigma_R(t))}{d(t - \sigma_R(t))}} \quad (8)$$

Note that the difference between (7) and (8) lies in the entry time. For (7), the entry time is t (which is regarded as current time) while that for (8) is $t - \sigma_R(t)$. If we regard $\tilde{t} = t - \sigma_R(t)$ as the current time, then t in (8) can be regarded as $t = \tilde{t} + h_R(n_R(\tilde{t}))$. (8) can be recognized in the same form as (7).

In this paper, we adopt the travel time function to define the flow conservation for the MFD system. We enforce the FIFO condition here to guarantee nonnegative flows.

Remark 2.1: The FIFO is generally required in the DTA literature for investigating route and departure time choices. Note that the FIFO principle can not be always fulfilled and its violation is permitted through overtaking in the real world, because vehicles can be given different characteristics even though they entered a region at the same time. In macroscopic (or aggregate) models vehicles are considered to take the same travel time to traverse a link (or region in our case) if they enter it at the same time. This is because macroscopic (or aggregate) models describe the average (i.e., statistical) behavior of vehicles. On the other hand, when there is a departure time choice (or assignment), the FIFO principle should be enforced for the issue of equity. Otherwise, drivers (be of the same route connecting the same origin-destination (OD) pair) depart earlier from the same origin will arrive at their destination later, i.e., the FIFO is violated. As a consequence, no driver would follow the departure time assignment to minimize the system cost given by the DSO equilibrium. The DSO equilibrium would be meaningless in this sense.

To enable route choice in the multi-region MFD system, an urban network is divided into several connected regions as depicted in Fig. 2(a). Fig. 2(b) shows an example of a path connecting the origin R_1 and the destination R_7 . The path flow would bypass two regions R_5 and R_6 successively before reaching the destination, i.e.,

$$\mathcal{R}^p = (R_1, R_5, R_6, R_7) \quad (9)$$

where \mathcal{R}^p denotes the set of all regions along path p . The network and its interconnection are described by a directed

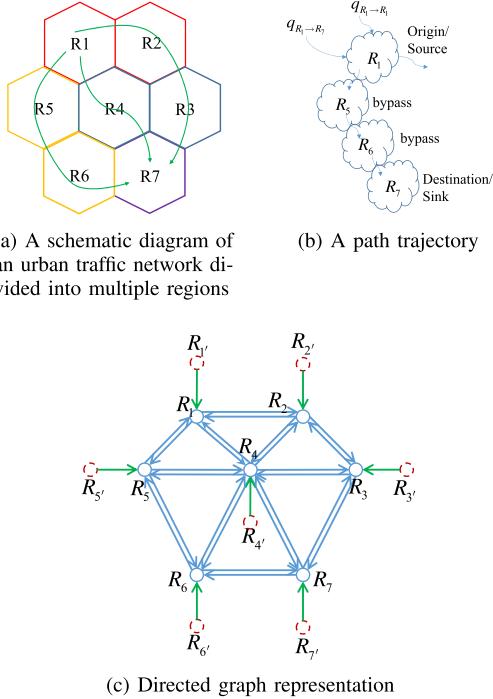


Fig. 2. Multi-region MFD system and the directed graph representation.

graph as depicted in Fig. 2(c). For travel demand whose origin and destination are in different regions, it is assumed that the flow is generated from a dummy source $R_{i'}$ and terminates in region R_j , $i \neq j$, where $R_{i'}$ denotes the dummy source and R_i the sink in region R_i . Separating the source and sink within the same region is important since a trip can be started and ended in the same region with a travel time equal to the average travel time of the region under the MFD framework while this is prohibited in the conventional directed graph representation of traffic network.

By the directed graph representation, we can define the sequential movements along different MFD regions using the concept of path/route by a sequence of connected regions, e.g.,

$$\mathcal{R}^p \doteq (R_0^p, R_1^p, R_2^p, \dots, R_i^p, \dots, R_{m(p)}^p), p \in \mathcal{P}$$

where R_0^p refers to the dummy source of path p , i.e., the origin; $m(p)$ is the number of regions along path p ; \mathcal{P} denotes the set of all paths in the network. For example, using this notation, (9) can be recast as

$$\mathcal{R}^p = (R_0^p, R_1^p, R_2^p, R_3^p, R_4^p) \doteq (R_{i'}, R_1, R_5, R_6, R_7)$$

To enable route choice and departure time choice, let $q^p(t)$ denote the departure rate to a particular path p at a particular time instant t . Along the path, flow exiting from the preceding upstream region is taken as input to the downstream region. For example, $G_{R_{i-1}^p}^p(t)$, the flow exiting from upstream region R_{i-1}^p by path p is regarded as the inflow to the downstream region R_i^p by path p , i.e., $q_{R_{i-1}^p}^p(t) = G_{R_{i-1}^p}^p(t)$. This flow propagation mechanism proceeds until this amount of flow arrives at the destination region $R_{m(p)}^p$, then we have the

following flow propagation dynamics for the regions along path p :

$$\frac{dn_{R_i^p}^p(t)}{dt} = q^p(t) - G_{R_i^p}^p(t), \quad (10a)$$

$$\frac{dn_{R_i^p}^p(t)}{dt} = G_{R_{i-1}^p}^p(t) - G_{R_i^p}^p(t), \quad \forall i \in [2, m(p)] \quad (10b)$$

where $n_{R_i^p}^p(t)$ denotes the network accumulation state of region R_i^p contributed by path p at time t . $m(p)$ denotes the number of regions along path p . As aforementioned, each region admits a time delay/travel time function $h_{R_i^p}(n_{R_i^p}(t))$, which defines the average travel time required to traverse region R_i^p , i.e.,

$$h_{R_i^p}(n_{R_i^p}(t)) = \frac{n_{R_i^p}(t)}{\mathcal{G}_{R_i^p}(n_{R_i^p}(t))} \quad (11)$$

where $n_{R_i^p}(t)$ denotes the total accumulation state of region R_i^p (the i^{th} region along path p) at time t . For a region R_j , the accumulation state $n_{R_j}(t)$ is given as:

$$n_{R_j}(t) = \sum_{p \in \mathcal{P}} \sum_{i=1}^{m(p)} n_{R_i^p}^p(t) \delta_{R_j, R_i^p} \quad (12)$$

As defined in Table I, the relation between the elements of \mathcal{R}^p and \mathcal{R} is as follows: If the i^{th} region along path p , i.e., the ‘virtual’ region R_i^p , is the ‘physical’ region region $R_j \in \mathcal{R}$, then $\delta_{R_j, R_i^p} = 1$. Otherwise, $\delta_{R_j, R_i^p} = 0$. $\sum_{i=1}^{m(p)}$ is used to determine whether path p traverses ‘physical’ region R_j . Note that there is at most one ‘virtual’ region along path p is region R_j . $\sum_{p \in \mathcal{P}}$ is used to evaluate the vehicle accumulation contributed by all paths.

We define $\tau_{R_i^p}^p(t)$ the exit time from the i^{th} region along path p , R_i^p , for vehicles departing from the origin at time t . To simplify the notation in deriving the DSO equilibrium condition, we use $\tau_{R_i^p}^p$ to replace $\tau_{R_i^p}^p(t)$, and denote the departure time as $\tau_{R_0^p}^p = t$. For a given departure flow profile $q^p(t)$ of path p at time t , the travel time for traversing the first region R_1^p can be calculated by (11). Combining (7), the exit time $\tau_{R_1^p}^p(t)$ and path-based flow rate $G_{R_1^p}^p$ exiting region R_1^p can be obtained. Proceeding along this stream, for all region ($i = 1, 2, \dots, m(p)$) along path p , the regional travel time and flow propagation can be represented as:

$$\tau_{R_i^p}^p(t) = \tau_{R_{i-1}^p}^p(t) + h_{R_i^p}(n_{R_i^p}(\tau_{R_{i-1}^p}^p)) \quad (13a)$$

$$h_{R_i^p}(n_{R_i^p}(\tau_{R_{i-1}^p}^p)) = \frac{n_{R_i^p}(\tau_{R_{i-1}^p}^p)}{\mathcal{G}_{R_i^p}(n_{R_i^p}(\tau_{R_{i-1}^p}^p))} \quad (13b)$$

$$G_{R_i^p}^p(\tau_{R_i^p}^p) = \frac{G_{R_{i-1}^p}^p(\tau_{R_{i-1}^p}^p)}{1 + h'_{R_i^p}(n_{R_i^p}(\tau_{R_{i-1}^p}^p)) \dot{n}_{R_i^p}(\tau_{R_{i-1}^p}^p)} \quad (13c)$$

According to the definition, the instantaneous journey time of a specific path p departing at time t is evaluated as

$$h^p(t, \mathbf{n}) = \sum_{i=1}^{m(p)} h_{R_i^p}(n_{R_i^p}(t)) \quad (14)$$

where $\mathbf{n} = (n_{R_j} : \forall R_j \in \mathcal{R})$.

To enforce departure time choice, a schedule delay cost function defining the early/late arrival penalty, say $\kappa(\cdot)$, is required. Combining this penalty with the journey time achieves the effective path delay/cost operators, i.e.,

$$\begin{aligned} \Psi^p(t, \mathbf{n}) &= h^p(t, \mathbf{n}) + \kappa^p(t, h^p(t, \mathbf{n})) \\ &= \sum_{i=1}^{m(p)} h_{R_i^p}(n_{R_i^p}(t)) + \kappa^p\left(t, \sum_{i=1}^{m(p)} h_{R_i^p}(n_{R_i^p}(t))\right) \end{aligned} \quad (15)$$

where $h^p(t, \mathbf{n})$ is a function of accumulative state \mathbf{n} . And $\kappa^p(t, h^p(t, \mathbf{n}))$ is a function of the difference between the actual arrival time $t + h^p(t, \mathbf{n})$ and the preferred arrival time t^* , i.e., $\chi = t + h^p(t, \mathbf{n}) - t^*$, with $t^* < T$ and the preferred arrival time t^* can be OD-dependent.

III. OPTIMAL CONTROL FORMULATION AND EQUILIBRIUM CONDITION FOR THE DYNAMIC SYSTEM OPTIMUM

For a set of connected regions, consider a finite time planning horizon $t \in [0, T]$, $T > 0$, each OD pair $w \in \mathcal{W}$ is with a fixed total amount of demand Q_w to be served. It is assumed that there is a central traffic manager to coordinate all commuters' departure times and routes during a certain time period to minimize the total system cost to achieve a DSO equilibrium. Considering the dynamic nature, we formulate the DSO problem as an optimal control problem. It seeks optimal path departure rate $q^{p*}(t)$, $\forall p \in \mathcal{P}$ to minimize the total system travel cost within the study period $[0, T]$, given a fixed total amount of traffic Q_w to be served between each origin-destination (OD) pair $w \in \mathcal{W}$. We define the DSO problem when the path travel time is evaluated by the instantaneous journey time as DSO:

$$\begin{aligned} \min J &= \sum_{w \in \mathcal{W}} \sum_{p \in \mathcal{P}_w} \int_0^T \Psi^p(t, \mathbf{n}) q^p(t) dt \\ &= \sum_{p \in \mathcal{P}} \int_0^T \Psi^p(t, \mathbf{n}) q^p(t) dt \end{aligned}$$

subject to

$$\frac{dn_{R_1^p}(t)}{dt} = q^p(t) - G_{R_1^p}(t), \quad (\alpha_{R_1^p}) \quad (16a)$$

$$\begin{aligned} \frac{dn_{R_i^p}(t)}{dt} &= G_{R_{i-1}^p}(t) - G_{R_i^p}(t), \quad (\alpha_{R_i^p}) \\ &\times \left(1 + h'_{R_i^p}(n_{R_i^p}(t)) \cdot \dot{n}_{R_i^p}(t)\right) \end{aligned} \quad (16b)$$

$$= q^p(t), \quad (\beta_{R_1^p}) \quad (16c)$$

$$\begin{aligned} &\times \left(1 + h'_{R_i^p}(n_{R_i^p}(t)) \cdot \dot{n}_{R_i^p}(t)\right) \\ &\times \left(1 + h'_{R_i^p}(n_{R_i^p}(t)) \cdot \dot{n}_{R_i^p}(t)\right) \end{aligned} \quad (16d)$$

$$-q^p(t) \leq 0, \quad (\gamma_{R_0^p}) \quad (16e)$$

$$-G_{R_i^p}(t) \leq 0, \quad (\gamma_{R_i^p}) \quad (16f)$$

$$-n_{R_i^p}(t) \leq 0, \quad (\lambda_{R_i^p}) \quad (16g)$$

$$\frac{dE_w(t)}{dt} = \sum_{p \in \mathcal{P}_w} q^p(t), \quad (\rho_w) \quad \text{总的输入?} \quad (16h)$$

$$E_w(0) = 0, \quad E_w(T) = Q_w, \quad (\phi_w) \quad (16i)$$

$$\begin{aligned} &\sum_{p \in \mathcal{P}} \left[q^p(t) \delta_{R_j, R_1^p} + \sum_{i=2}^{m(p)} G_{R_{i-1}^p}(t) \delta_{R_j, R_i^p} \right] \\ &\leq q_{R_j}^{\max}, \quad (\zeta_{R_j}) \end{aligned} \quad (16j)$$

where (16a)-(16i) are defined for all $w \in \mathcal{W}$, $p \in \mathcal{P}_w$, $i \in [2, m(p)]$, (16j) is defined for all $R_j \in \mathcal{R}$, respectively. $n_{R_j}(t)$ and $n_{R_i^p}(t)$ denote the total accumulation state of region R_j and the total accumulation state of 'virtual' region R_i^p , the i^{th} region along path p , at time t , respectively; $n_{R_i^p}(t)$ denotes the accumulation of region R_i^p contributed by path p at time t . ‘’’ denotes differentiation with respect to the associated function argument,³ while ‘’’ denotes differentiation with respect to time t . \mathcal{P}_w is the set of paths connecting OD pair w and $E_w(t)$ is an extended state for flow conservation of OD pair w . δ_{R_j, R_i^p} is the Kronecker Delta function. If the i^{th} region along path p , i.e., 'virtual' region R_i^p , is the 'physical' region R_j ($R_j \in \mathcal{R}$), then $\delta_{R_j, R_i^p} = 1$; otherwise, $\delta_{R_j, R_i^p} = 0$. The control variables are $\mathbf{q} = (q^p : p \in \mathcal{P})$, $\mathbf{G} = (G_{R_i^p}^p : p \in \mathcal{P}, i \in [1, m(p)])$, where $G_{R_i^p}^p$ denotes the outflow rate exiting region R_i^p along path p . (16a)-(16b) are network traffic dynamics corresponding to (10a)-(10b). (16c)-(16d) are flow propagation constraints, which is an equivalent version of (7). Note for any given $q^p(t) \geq 0$, $\forall p \in \mathcal{P}$, we have $G_{R_j}^p(t) \geq 0$, $\forall p \in \mathcal{P}$, $\forall R_j \in \mathcal{R}$,⁴ and $n_{R_j}(t) \geq 0$, $\forall R_j \in \mathcal{R}$ under FIFO condition. (16h) and (16i) are flow conservation constraints. Without loss of generality, it is assumed that the network is empty at the beginning, i.e., $n_{R_j}(0) = 0, \forall R_j \in \mathcal{R}$.⁵ (16j) restricts that the inflow to a region should be less than its maximum service rate or capacity, and $q_{R_j}^{\max}$ denote the maximum inflow rate of R_j . Variables in brackets

³It is assumed that all network delay functions are differentiable with regard to their own arguments.

⁴Given the saturation constraint (16e) holds, we have $G_{R_i^p}^p(t) \geq 0$ since MFD establishes a non-negative mapping from accumulation state n_i to the outflow rate.

⁵This is a mathematical convenience widely adopted in the DTA literature. From mathematical point of view, nonzero initial condition can be also assumed without changing the analysis but the boundary condition.

of (16a)-(16i) are Lagrange multipliers associated with the corresponding constraints, respectively.

Proposition 3.1: *The equilibrium condition for Problem **DSO** with simultaneous route and departure time assignments can be stated as follows:*

$$q^p(t) \begin{cases} > 0 \Rightarrow \Psi^p(t, \mathbf{n}^*) + \varphi^p(t) + l^p(t) = \phi_w, & \forall w \in \mathcal{W}, p \in \mathcal{P}_w \\ = 0 \Rightarrow \Psi^p(t, \mathbf{n}^*) + \varphi^p(t) + l^p(t) > \phi_w, \end{cases}$$

where \mathbf{n}^* is the vector of optimal network accumulations, ϕ_w specifies the theoretical OD cost of the OD pair w , i.e., the minimum (or equilibrium) travel cost of the OD pair w . The generalized travel cost consists of three parts, effective path cost $\Psi^p(t, \mathbf{n}^*)$, dynamic marginal cost $\varphi^p(t)$, and dynamic external cost $l^p(t)$. The equilibrium condition implies that, at any time instant t , the generalized travel cost on any path with positive flow, i.e., $q^p(t) > 0$, must equal a constant, ϕ_w .

$$\varphi^p(t) = \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} \left[q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} \right] \Big|_{\mathbf{n}=\mathbf{n}^*} ds \quad (17)$$

$$l^p(t) = \sum_{i=1}^{m(p)} \zeta_{R_i^p} \left(\tau_{R_{i-1}^p}^p \right) \quad (18)$$

where, (19), shown at the bottom of this page, $\varphi^p(t)$ denotes the dynamic marginal cost of path p , which represents the sensitivity value of the total system travel cost with respect to the perturbations of regional accumulations contributed by the path inflow $q^p(t)$ at time t . For a specific region R_i^p along the path p , this perturbation will become effective at the time when the flow of $q^p(t)$ enters the region while will become ineffective at the time when the flow exits from the region, i.e., the effective time interval is $[\tau_{R_{i-1}^p}^p, \tau_{R_i^p}^p]$. Note that all the paths traversing the region will be affected by this perturbation. Therefore, we sum the sensitivity value of all the paths traversing the region during effective time interval. For a path p' , this path is affected by the perturbation of region R_i^p when path p' shares region R_i^p . As shown in (19), when the region R_i^p is part of path p' , the marginal cost of path p' caused by the variation of travel time due to the perturbation

of region R_i^p is nonzero. Otherwise, the marginal cost will be zero.

$l^p(t)$ represents the dynamic external cost imposed on travelers caused by their presence on the region along the trip trajectory, wherein $\zeta_{R_i^p}$ denotes the Lagrange multiplier associated with the inflow saturation constraint imposed on R_i^p , and $\tau_{R_{i-1}^p}^p$ denotes the entry time to R_i^p of vehicles departing from the origin at time t .

Proof: In this proof, the equilibrium condition for the Problem **DSO** is derived by extending the Pontryagin Minimum Principle for the DUE problem developed in [21], [49], [50]. To simplify the notation, we define

$$G_{R_0^p}^p(t) = q^p(t), \quad \forall p \in \mathcal{P}$$

$$\tilde{G}_{R_i^p}^p(t) = G_{R_i^p}^p \left(t + h_{R_i^p} \left(n_{R_i^p}(t) \right) \right), \quad \forall p \in \mathcal{P}, i \in [1, m(p)]$$

太复杂没看懂。。。

To apply the Pontryagin Minimum Principle, the Hamiltonian function for the optimal control problem is defined in (20), shown at the bottom of the next page. To analyze the first-order optimality, we derive the partial derivatives of the Hamiltonian function in (21a)-(21e), shown at the bottom of the next page.

Stationary conditions at optimality are

$$\frac{\partial H}{\partial q^p} \Big|_t = 0, \quad \forall p \in \mathcal{P} \quad (22a)$$

$$\begin{aligned} \frac{\partial H}{\partial G_{R_i^p}^p} \Big|_{\tau_{R_i^p}^p} + \left[\frac{\partial H}{\partial \tilde{G}_{R_i^p}^p} \frac{1}{1 + h'_{R_i^p}(n_{R_i^p}) \cdot \dot{n}_{R_i^p}} \right] \Big|_{\tau_{R_{i-1}^p}^p} \\ = 0, \end{aligned} \quad \forall p \in \mathcal{P}, i \in [1, m(p)] \quad (22b)$$

The adjoint equations are

$$\frac{d\alpha_{R_i^p}^p(t)}{dt} = -\frac{\partial H(t)}{\partial n_{R_i^p}^p(t)}, \quad \forall p \in \mathcal{P}, i \in [1, m(p)] \quad (23a)$$

$$\frac{d\rho_w(t)}{dt} = 0, \quad \forall w \in \mathcal{W} \quad (23b)$$

$$\begin{aligned} \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} &= \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial h^{p'}(s, \mathbf{n})} \frac{\partial h^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} \\ &= \left[1 + \frac{\partial \kappa^{p'}(s, h^{p'}(s, \mathbf{n}))}{\partial h^{p'}(s, \mathbf{n})} \right] \left[\sum_{i'=1}^{m(p')} \frac{\partial h_{R_{i'}^{p'}}(n_{R_{i'}^{p'}}(s))}{\partial n_{R_i^p}(s)} \right] \\ &= \begin{cases} \left[1 + \frac{\partial \kappa^{p'}(s, h^{p'}(s, \mathbf{n}))}{\partial h^{p'}(s, \mathbf{n})} \right] \frac{dh_{R_i^p}(n_{R_i^p}(s))}{dn_{R_i^p}(s)}, & \text{if } R_i^p \in \mathcal{R}^{p'} \\ 0, & \text{otherwise} \end{cases} \\ &= \left[1 + \frac{\partial \kappa^{p'}(s, h^{p'}(s, \mathbf{n}))}{\partial h^{p'}(s, \mathbf{n})} \right] \frac{dh_{R_i^p}(n_{R_i^p}(s))}{dn_{R_i^p}(s)} \sum_{i'=1}^{m(p')} \delta_{R_i^p, R_{i'}^{p'}} \end{aligned} \quad (19)$$

Boundary conditions at terminal time T are

$$-\rho_w(T) - \phi_w = 0, \forall w \in \mathcal{W} \quad (24)$$

Complementary slackness conditions for nonnegative flow constraints and saturation constraints (16e)-(16g) are:

$$q^p(t) \geq 0, \gamma_{R_0^p}^p(t) \geq 0, q^p(t)\gamma_{R_0^p}^p(t) = 0 \quad (25a)$$

$$G_{R_i^p}^p(t) \geq 0, \gamma_{R_i^p}^p(t) \geq 0, G_{R_i^p}^p(t)\gamma_{R_i^p}^p(t) = 0 \quad (25b)$$

$$n_{R_i^p}^p(t) \geq 0, \lambda_{R_i^p}^p(t) \geq 0, n_{R_i^p}^p(t)\lambda_{R_i^p}^p(t) = 0 \quad (25c)$$

for all $p \in \mathcal{P}, i \in [1, m(p)]$.

In conjunction with (21a), (22a) gives

$$\begin{aligned} \frac{\partial H}{\partial q^p} \Big|_t &= \Psi^p(t, \mathbf{n}) + \alpha_{R_1^p}^p(t) + \beta_{R_1^p}^p(t) - \gamma_{R_0^p}^p(t) + \rho_w + \zeta_{R_1^p}^p(t) \\ &= 0, \forall w \in \mathcal{W}, p \in \mathcal{P}_w \end{aligned} \quad (26)$$

In conjunction with (21e), (23a) gives

$$\frac{d\alpha_{R_i^p}^p(t)}{dt} = -\frac{\partial H(t)}{\partial n_{R_i^p}^p(t)} = -\sum_{p' \in \mathcal{P}} q^{p'}(t) \frac{\partial \Psi^{p'}(t, \mathbf{n})}{\partial n_{R_i^p}^p(t)} + \lambda_{R_i^p}^p(t) \quad (27)$$

for all $p \in \mathcal{P}, i \in [1, m(p)]$.

Hence

$$\begin{aligned} &\alpha_{R_i^p}^p(\tau_{R_{i-1}^p}^p) - \alpha_{R_i^p}^p(\tau_{R_i^p}^p) \\ &= \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}^p(s)} ds - \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds \end{aligned} \quad (28)$$

for all $p \in \mathcal{P}, i \in [1, m(p)]$.

Using (21b), (21c), (21d), (22b) gives

$$\begin{aligned} &-\alpha_{R_i^p}^p\left(\tau_{R_i^p}^p\right) + \alpha_{R_{i+1}^p}^p\left(\tau_{R_i^p}^p\right) + \beta_{R_{i+1}^p}^p\left(\tau_{R_i^p}^p\right) - \gamma_{R_i^p}^p\left(\tau_{R_i^p}^p\right) \\ &+ \zeta_{R_{i+1}^p}\left(\tau_{R_i^p}^p\right) - \beta_{R_i^p}^p\left(\tau_{R_{i-1}^p}^p\right) = 0, i \in [1, m(p)-1] \end{aligned} \quad (29a)$$

$$-\alpha_{R_{m(p)}}^p\left(\tau_{R_{m(p)}}^p\right) - \gamma_{R_{m(p)}}^p\left(\tau_{R_{m(p)}}^p\right) - \beta_{R_{m(p)}}^p\left(\tau_{R_{m(p)-1}}^p\right) = 0 \quad (29b)$$

for all $p \in \mathcal{P}$. So that

$$\begin{aligned} &\alpha_{R_{i+1}^p}^p\left(\tau_{R_i^p}^p\right) + \beta_{R_{i+1}^p}^p\left(\tau_{R_i^p}^p\right) \\ &= \alpha_{R_i^p}^p\left(\tau_{R_i^p}^p\right) + \beta_{R_i^p}^p\left(\tau_{R_{i-1}^p}^p\right) - \zeta_{R_{i+1}^p}\left(\tau_{R_i^p}^p\right) + \gamma_{R_i^p}^p\left(\tau_{R_i^p}^p\right) \end{aligned} \quad (30)$$

$$\begin{aligned} H_1(t) &= \sum_{p \in \mathcal{P}} \Psi_1^p(t, \mathbf{n}) G_{R_0^p}^p(t) + \sum_{p \in \mathcal{P}} \sum_{i=1}^{m(p)} \alpha_{R_i^p}^p(t) \left[G_{R_{i-1}^p}^p(t) - G_{R_i^p}^p(t) \right] \\ &+ \sum_{p \in \mathcal{P}} \sum_{i=1}^{m(p)} \beta_{R_i^p}^p(t) \left\{ G_{R_{i-1}^p}^p(t) - \tilde{G}_{R_i^p}^p(t) \left[1 + h'_{R_i^p}(n_{R_i^p}(t)) \cdot \dot{n}_{R_i^p}(t) \right] \right\} - \sum_{p \in \mathcal{P}} \left[\sum_{i=0}^{m(p)} \gamma_{R_i^p}^p(t) G_{R_i^p}^p(t) + \sum_{i=1}^{m(p)} \lambda_{R_i^p}^p(t) n_{R_i^p}^p(t) \right] \\ &+ \sum_{w \in \mathcal{W}} \rho_w(t) \left(\sum_{p \in \mathcal{P}_w} G_{R_0^p}^p(t) \right) + \sum_{R_j \in \mathcal{R}} \zeta_{R_j}(t) \left[\sum_{p \in \mathcal{P}} \sum_{i=1}^{m(p)} G_{R_{i-1}^p}^p(t) \delta_{R_j, R_i^p} - q_{R_j}^{\max} \right] \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{\partial H(t)}{\partial q^p(t)} &= \frac{\partial H(t)}{\partial G_{R_0^p}^p(t)} = \Psi^p(t, \mathbf{n}) + \alpha_{R_1^p}^p(t) + \beta_{R_1^p}^p(t) - \gamma_{R_0^p}^p(t) + \rho_w(t) + \sum_{R_j \in \mathcal{R}} \zeta_{R_j}(t) \delta_{R_j, R_1^p} \\ &= \Psi^p(t, \mathbf{n}) + \alpha_{R_1^p}^p(t) + \beta_{R_1^p}^p(t) - \gamma_{R_0^p}^p(t) + \rho_w(t) + \zeta_{R_1^p}(t), \forall p \in \mathcal{P} \end{aligned} \quad (21a)$$

$$\begin{aligned} \frac{\partial H(t)}{\partial G_{R_i^p}^p(t)} &= -\alpha_{R_i^p}^p(t) + \alpha_{R_{i+1}^p}^p(t) + \beta_{R_{i+1}^p}^p(t) - \gamma_{R_i^p}^p(t) + \sum_{R_j \in \mathcal{R}} \zeta_{R_j}(t) \delta_{R_j, R_{i+1}^p} \\ &= -\alpha_{R_i^p}^p(t) + \alpha_{R_{i+1}^p}^p(t) + \beta_{R_{i+1}^p}^p(t) - \gamma_{R_i^p}^p(t) + \zeta_{R_{i+1}^p}(t), \forall p \in \mathcal{P}, i \in [1, m(p)-1] \end{aligned} \quad (21b)$$

$$\frac{\partial H(t)}{\partial G_{R_{m(p)}}^p(t)} = -\alpha_{R_{m(p)}}^p(t) - \gamma_{R_{m(p)}}^p(t), \forall p \in \mathcal{P} \quad (21c)$$

$$\frac{\partial H(t)}{\partial \tilde{G}_{R_i^p}^p(t)} = -\beta_{R_i^p}^p(t) \left[1 + h'_{R_i^p}(n_{R_i^p}(t)) \cdot \dot{n}_{R_i^p}(t) \right], \forall p \in \mathcal{P}, i \in [1, m(p)] \quad (21d)$$

$$\frac{\partial H(t)}{\partial n_{R_i^p}^p(t)} = \sum_{p' \in \mathcal{P}} q^{p'}(t) \frac{\partial \Psi^{p'}(t, \mathbf{n})}{\partial n_{R_i^p}^p(t)} + \lambda_{R_i^p}^p(t) = \sum_{p' \in \mathcal{P}} q^{p'}(t) \frac{\partial \Psi^{p'}(t, \mathbf{n})}{\partial n_{R_i^p}^p(t)} - \lambda_{R_i^p}^p(t), \forall p \in \mathcal{P}, i \in [1, m(p)] \quad (21e)$$

for all $p \in \mathcal{P}$, $i \in [1, m(p) - 1]$, and

$$\beta_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)-1}^p}^p \right) = -\alpha_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)}^p}^p \right) - \gamma_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)}^p}^p \right) \quad (31)$$

for all $p \in \mathcal{P}$.

Define $D_i^p = \alpha_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) + \beta_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right)$, $i \in [1, m(p)]$. From (28) and (31), we have

$$\begin{aligned} D_{m(p)}^p &= \alpha_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)-1}^p}^p \right) + \beta_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)-1}^p}^p \right) \\ &= \alpha_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)-1}^p}^p \right) - \alpha_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)}^p}^p \right) \\ &\quad - \gamma_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)}^p}^p \right) \\ &= \int_{\tau_{R_{m(p)-1}^p}^p}^{\tau_{R_{m(p)}^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_{m(p)}^p}(s)} ds \\ &\quad - \int_{\tau_{R_{m(p)-1}^p}^p}^{\tau_{R_{m(p)}^p}^p} \lambda_{R_{m(p)}^p}^p(s) ds \\ &\quad - \gamma_{R_{m(p)}^p}^p \left(\tau_{R_{m(p)}^p}^p \right) \end{aligned}$$

Using (28) and (30), $\forall i \in [1, m(p) - 1]$, we have

$$\begin{aligned} D_{i+1}^p &= \alpha_{R_{i+1}^p}^p \left(\tau_{R_i^p}^p \right) + \beta_{R_{i+1}^p}^p \left(\tau_{R_i^p}^p \right) \\ &= \alpha_{R_i^p}^p \left(\tau_{R_i^p}^p \right) + \beta_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) \\ &\quad - \zeta_{R_{i+1}^p}^p \left(\tau_{R_i^p}^p \right) + \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) \\ D_i^p &= \alpha_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) + \beta_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) \end{aligned}$$

Further we have

$$\begin{aligned} D_{i+1}^p - D_i^p &= \alpha_{R_i^p}^p \left(\tau_{R_i^p}^p \right) - \alpha_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) - \zeta_{R_{i+1}^p}^p \left(\tau_{R_i^p}^p \right) + \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) \\ &= - \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} ds + \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds \\ &\quad - \zeta_{R_{i+1}^p}^p \left(\tau_{R_i^p}^p \right) + \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) \end{aligned}$$

Then we have

$$D_1^p = - \sum_{i=1}^{m(p)-1} [D_{i+1}^p - D_i^p] + D_{m(p)}^p$$

Thus

$$\begin{aligned} D_1^p &= \sum_{i=1}^{m(p)-1} \left[\int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} ds + \zeta_{R_{i+1}^p}^p \left(\tau_{R_i^p}^p \right) \right] \\ &\quad - \sum_{i=1}^{m(p)-1} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds - \sum_{i=1}^{m(p)-1} \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) + D_{m(p)}^p \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} ds + \sum_{i=2}^{m(p)} \zeta_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) \\ &\quad - \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds - \sum_{i=1}^{m(p)} \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) \\ \text{Since } D_1^p &= \alpha_{R_1^p}^p \left(\tau_{R_0^p}^p \right) + \beta_{R_1^p}^p \left(\tau_{R_0^p}^p \right), \text{ we have} \\ \alpha_{R_1^p}^p(t) + \beta_{R_1^p}^p(t) &- \gamma_{R_0^p}^p(t) + \zeta_{R_1^p}^p(t) \\ &= D_1^p - \gamma_{R_0^p}^p(t) + \zeta_{R_1^p}^p(t) \\ &= \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} ds + \sum_{i=1}^{m(p)} \zeta_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) \\ &\quad - \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds - \sum_{i=0}^{m(p)} \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) \end{aligned} \quad (32)$$

From (23b) and (24), we have

$$\rho_w(t) = \rho_w(T) = -\phi_w, \quad \forall w \in \mathcal{W} \quad (33)$$

Then (26) give

$$\begin{aligned} \Psi^p(t, \mathbf{n}) + \alpha_{R_1^p}^p(t) + \beta_{R_1^p}^p(t) - \lambda_{R_0^p}^p(t) + \zeta_{R_1^p}^p(t) &= \Psi^p(t, \mathbf{n}) + \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} ds \\ &\quad + \sum_{i=1}^{m(p)} \zeta_{R_i^p}^p \left(\tau_{R_{i-1}^p}^p \right) - \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds - \sum_{i=0}^{m(p)} \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) \\ &= \phi_w, \quad \forall w \in \mathcal{W}, p \in \mathcal{P}_w \end{aligned} \quad (34)$$

Let path p with $\mathcal{R}^p = (R_1^p, \dots, R_i^p, \dots, R_{m(p)}^p)$ is a path chosen (also known as open path) at time t (i.e., the departure time), i.e.

$$G_{R_0^p}^p(t) = q^p(t) > 0$$

By proper flow propagation and the non-negative constraint, we have

$$G_{R_i^p}^p(\tau_{R_i^p}^p) > 0, \quad \forall i \in [1, m(p)]$$

$$n_{R_i^p}^p(s) > 0, \quad \forall s \in [\tau_{R_{i-1}^p}^p, \tau_{R_i^p}^p], \quad i \in [1, m(p)]$$

According to (25a)-(25c), the Lagrange multipliers of the complementary slackness conditions are

$$\gamma_{R_0^p}^p(\tau_{R_0^p}^p) = 0 \quad (35a)$$

$$\gamma_{R_i^p}^p(\tau_{R_i^p}^p) = 0, \quad \forall i \in [1, m(p)] \quad (35b)$$

$$\lambda_{R_i^p}^p(s) = 0, \quad \forall s \in [\tau_{R_{i-1}^p}^p, \tau_{R_i^p}^p], \quad i \in [1, m(p)] \quad (35c)$$

If $q^p(t) > 0$, we have

$$\sum_{i=0}^{m(p)} \gamma_{R_i^p}^p \left(\tau_{R_i^p}^p \right) = 0, \quad \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}^p}^p}^{\tau_{R_i^p}^p} \lambda_{R_i^p}^p(s) ds = 0$$

Define:

$$\begin{aligned}\varphi^p(t) &= \sum_{i=1}^{m(p)} \int_{\tau_{R_{i-1}}^p}^{\tau_i^p} \sum_{p' \in \mathcal{P}} q^{p'}(s) \frac{\partial \Psi^{p'}(s, \mathbf{n})}{\partial n_{R_i^p}(s)} ds \\ l^p(t) &= \sum_{i=1}^{m(p)} \zeta_{R_i^p} \left(\tau_{R_{i-1}}^p \right)\end{aligned}$$

Then from (34), the DSO condition with saturation constraints can be cast as: for an open path $p \in \mathcal{P}_w$, $q^{p*}(t) > 0$, we have the following equilibrium condition

$$\Psi^p(t, \mathbf{n}^*) + \varphi^p(t) + l^p(t) = \phi_w \Rightarrow q^{p*}(t) > 0$$

otherwise

$$\Psi^p(t, \mathbf{n}^*) + \varphi^p(t) + l^p(t) > \phi_w \Rightarrow q^{p*}(t) = 0$$

Then we can deduce the equilibrium condition in Proposition 3.1. ■

Remark 3.1: In this remark, we show that the DSO equilibrium condition in Proposition 3.1. can reduce to the conventional static SO condition when the link (or region) travel cost is separable. The static SO assumes that there is a central traffic manager to coordinate all commuters' path choices to minimize the network travel time, i.e.,

$$\begin{aligned}\min J(\mathbf{n}) &= \sum_{w \in \mathcal{W}} \sum_{p \in \mathcal{P}_w} \Psi^p(\mathbf{n}) q^p = \sum_{p \in \mathcal{P}} \Psi^p(\mathbf{n}) q^p \\ &= \sum_{R_j \in \mathcal{R}} h_{R_j}(n_{R_j}) n_{R_j}\end{aligned}$$

subject to

$$q^p = n_{R_i^p}, (\alpha_{R_i^p}), \forall p \in \mathcal{P}, i \in [1, m(p)] \quad (36a)$$

$$-q^p \leq 0, (\gamma_{R_0^p}), \forall p \in \mathcal{P} \quad (36b)$$

$$-n_{R_i^p} \leq 0, (\gamma_{R_i^p}), \forall p \in \mathcal{P}, i \in [1, m(p)] \quad (36c)$$

$$\sum_{p \in \mathcal{P}_w} q^p = Q_w, (\rho_w), \forall w \in \mathcal{W} \quad (36d)$$

where

$$\Psi^p(\mathbf{n}) = \sum_{i=1}^{m(p)} h_{R_i^p}(n_{R_i^p}), \forall p \in \mathcal{P} \quad (37a)$$

$$n_{R_j} = \sum_{p \in \mathcal{P}} \sum_{i=1}^{m(p)} n_{R_i^p} \delta_{R_j, R_i^p}, \forall R_j \in \mathcal{R} \quad (37b)$$

Lemma 3.1: (See Sheffi [51], Sec 3.4) The equilibrium condition for the static SO can be stated as follows:

$$q^p \begin{cases} > 0 \Rightarrow \Psi^p(\mathbf{n}^*) + \varphi^p = \phi_w, \forall w \in \mathcal{W}, p \in \mathcal{P}_w \\ = 0 \Rightarrow \Psi^p(\mathbf{n}^*) + \varphi^p > \phi_w, \end{cases} \quad (38)$$

where

$$\varphi^p = \sum_{i=1}^{m(p)} \sum_{p' \in \mathcal{P}} q^{p'} \frac{\partial \Psi^{p'}}{\partial n_{R_i^p}} = \sum_{i=1}^{m(p)} \sum_{p' \in \mathcal{P}} \left(q^{p'} \frac{dh_{R_i^p}^{p'}}{dn_{R_i^p}} \sum_{i'=1}^{m(p')} \delta_{R_i^p, R_{i'}^{p'}} \right) \quad (39)$$

For comparison, we have defined the total travel time in terms of path-based formulation in line with **DSO**. We prove in the online appendix that

$$J(\mathbf{n}) = \sum_{p \in \mathcal{P}} \Psi^p(\mathbf{n}) q^p = \sum_{R_j \in \mathcal{R}} h_{R_j}(n_{R_j}) n_{R_j} \quad (40)$$

We deduce (39) using the following facts:

$$\begin{aligned}\frac{\partial h_{R_i^p}(n_{R_i^p})}{\partial n_{R_i^p}} \\ = \begin{cases} \frac{dh_{R_i^p}(n_{R_i^p})}{dn_{R_i^p}}, & \text{if } R_i^p = R_{i'}^{p'} \\ 0, & \text{otherwise} \end{cases} = \frac{dh_{R_i^p}(n_{R_i^p})}{dn_{R_i^p}} \delta_{R_i^p, R_{i'}^{p'}}\end{aligned}$$

and

$$\sum_{R_j \in \mathcal{R}} \delta_{R_j, R_i^p} = 1$$

$$n_{R_i^p} = \sum_{p' \in \mathcal{P}} \sum_{i'=1}^{m(p')} n_{R_{i'}^{p'}} \delta_{R_i^p, R_{i'}^{p'}} = \sum_{p' \in \mathcal{P}} \sum_{i=1}^{m(p')} q^{p'} \delta_{R_i^p, R_{i'}^{p'}}$$

Then we have

$$\begin{aligned}\varphi^p &= \sum_{i=1}^{m(p)} \sum_{p' \in \mathcal{P}} q^{p'} \frac{\partial \Psi^{p'}}{\partial n_{R_i^p}} = \sum_{i=1}^{m(p)} \sum_{p' \in \mathcal{P}} q^{p'} \left(\sum_{i'=1}^{m(p')} \frac{\partial h_{R_{i'}^{p'}}(n_{R_{i'}^{p'}})}{\partial n_{R_i^p}} \right) \\ &= \sum_{i=1}^{m(p)} \sum_{p' \in \mathcal{P}} q^{p'} \left(\frac{dh_{R_i^p}(n_{R_i^p})}{dn_{R_i^p}} \sum_{i'=1}^{m(p')} \delta_{R_i^p, R_{i'}^{p'}} \right) \\ &= \sum_{i=1}^{m(p)} \sum_{p' \in \mathcal{P}} \sum_{i'=1}^{m(p')} q^{p'} \frac{dh_{R_i^p}(n_{R_i^p})}{dn_{R_i^p}} \delta_{R_i^p, R_{i'}^{p'}} \\ &= \sum_{i=1}^{m(p)} \frac{dh_{R_i^p}(n_{R_i^p})}{dn_{R_i^p}} \left(\sum_{p' \in \mathcal{P}} \sum_{i'=1}^{m(p')} q^{p'} \delta_{R_i^p, R_{i'}^{p'}} \right) \\ &= \sum_{i=1}^{m(p)} n_{R_i^p} \frac{dh_{R_i^p}(n_{R_i^p})}{dn_{R_i^p}}\end{aligned}$$

The definition of instantaneous journey time in **DSO** assumes the path travel time is additive and the region travel cost is separable. The same assumption is adopted in the conventional static SO as depicted in Lemma 3.1. Note that no capacity constraint is considered in the static SO. Therefore, $l^p(\cdot) = 0$ if we remove the inflow capacity constraint in the **DSO**. Since we do not consider time-dependent traffic condition in the static SO, there should be no integral in (17) under static network condition. Thus, (17) reduces to the form of (39). Furthermore, there is no departure time choice in the static SO and thus no early/late arrival penalty $\kappa(\cdot)$ should be considered. Therefore, $\frac{\partial \kappa}{\partial h} = 0$ in (19). With these observations, we can conclude the equilibrium condition in Proposition 3.1 reduces to the conventional SO condition in Lemma 3.1 for a static network without capacity constraint.

Remark 3.2: The travel time function is not well defined when $\mathcal{G}_R(n_R) = 0$ as $n_R = n_R^{jam}$. This phenomenon is recognized as the “gridlock” in the MFD literature that the

accumulation exceeding sustainable point leads to zero exit function [26], [52]. One of the objectives of traffic control is to prevent the network from over-saturation. In this paper, we adopt the inflow capacity constraint approach to achieve this goal. By imposing the inflow capacity constraint (16j), we can effectively avoid the “gridlock”. As a matter of fact, the inflow capacity constraint has been widely adopted in the MFD literature by tracing the flow propagation in the network loading numerically, see, e.g., [35]. The additional cost caused by the inflow capacity constraint can be regarded as the price to prevent the network entering over-saturated conditions.

IV. NUMERICAL EXAMPLES

In this section, the MFD of the form $a_1 \cdot n^3 - b_1 \cdot n^2 + c_1 \cdot n$ and the parameters as calibrated in the literature, e.g., [28], are adopted in the simulations.

A. Within-Region Trip

For the within region trips, travelers do not have route choice but departure time choice (i.e., the morning commute problem). Let us consider a single region MFD network with a nonlinear network travel time function as $h_R(n_R) = 1800/(1.4877 \cdot 10^{-7} \cdot n_R^2 - 2.9815 \cdot 10^{-3} \cdot n_R + 15.0912)$ (unit/time). It is assumed that the total amount of within region demand is $Q_{od} = 1500$ (units). The planning horizon is $T = 800$ (times) to ensure all travel demand can be served. The expected arrival time interval is $[t_{de}, t_{dl}] = [450, 550]$ (unit-time) with an early/late arrival penalty function defined as

$$\kappa^p(t, h^p) = \begin{cases} 0.1(t + h^p - t_{de})^2, & \text{if } t + h^p < t_{de} \\ 0, & \text{if } t_{de} \leq t + h^p \leq t_{dl} \\ 0.2(t + h^p - t_{dl})^2, & \text{if } t + h^p > t_{dl} \end{cases} \quad (41)$$

The upper bound of inflow rate is set as 12.6 (unit/time). For the within region trips, the system manager only needs to coordinate departure time for all travelers to minimize the total system cost throughout the planning horizon. All travelers would cooperatively choose their departure times according to the assignment.

Using the solution algorithm depicted in the online appendix, we obtain the following results for the DSO problem with departure time choice only. Without loss of generality, we assume zero initial conditions. Fig. 3 depicts the path departure rate against various travel costs over time. The ‘theoretical OD cost’ for a specific OD pair w is the minimum travel cost of the OD pair under the DSO equilibrium condition. It is the Lagrangian multiplier corresponding to the flow conservation of travel demand, i.e., ϕ_w . Comparing the path departure rate profile against the generalized travel cost profile, i.e., the blue line, in Fig. 3, at the first beginning, the generalized travel cost is larger than the equilibrium cost, no travelers would depart during the time interval. The generalized travel cost remains constant during the departure window that well satisfies the DSO condition. As indicated by Fig. 3, the dynamic external cost of the DSO case mainly consists of the dynamic marginal cost imposed on travelers

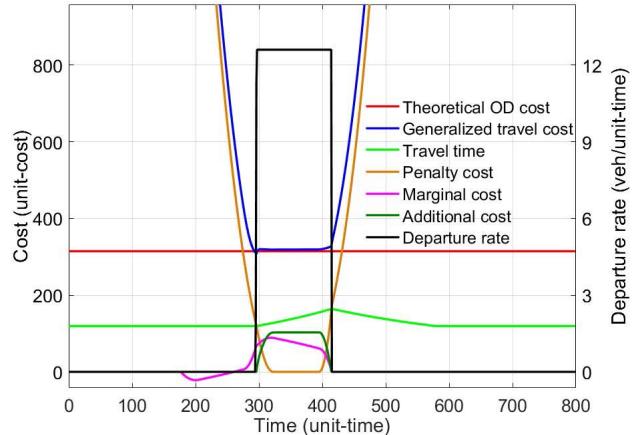


Fig. 3. Departure rate against travel costs under DSO equilibrium.

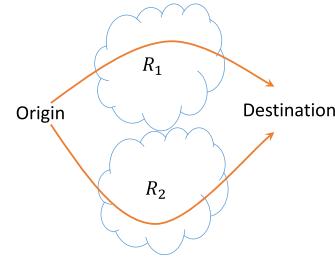


Fig. 4. An illustration of the network consists of two parallel regions.

by their presence on the region and the additional travel cost caused by the inflow constraint to the region, i.e., the capacity of the network. To avoid the marginal cost (or additional cost), travelers would expand the departure window to avoid network congestion. Such collaborative behavior helps minimize the total system cost. Similar findings can be found in the DSO analysis in the DTA literature [2], [5].

If tolling strategy is implemented, i.e., to charge the travelers a price equals to the difference in the generalized cost of the DUE and the DSO, to incentivize travelers to adjust their departure times (and thus their arrival times) to minimize the total cost of the system. Due to the early-late penalty function, the dynamic marginal cost can be also negative as illustrated in Fig. 3. Under this case, the “pricing” strategy can be also deployed by subsidizing the travelers to incentivize them to depart earlier to alleviate the peak-hour traffic congestion. As discussed in [25], the optimal pricing strategy can be implemented as a hybrid incentivizing strategy that prices a portion of the peak time and subsidizes the rest.

B. Parallel Regions

Fig. 4 depicts an OD pair that can be reached by crossing two parallel regions. Each region is regarded as a path in this example such that both route choice and departure time choice are enabled. The configuration of each region is specified in Table II. Region R_2 has a smaller free-flow time than region R_1 , while the travel time of region R_2 increases faster than that of region R_1 , and will exceed the travel time of region R_1 when the network accumulation reaches 500 units. The total OD travel demand is $Q_{od} = 2000$ (units). The planning horizon

TABLE II
NETWORK CONFIGURATION

Network configuration	Nonlinear network delay
R_1	$a_1 = 2 \cdot 1.4877 \cdot 10^{-7}$ $b_1 = 2 \cdot 2.9815 \cdot 10^{-3}$ $c_1 = 2 \cdot 15.0912$
R_2	$\frac{3600 \cdot n}{a_1 \cdot n^3 - b_1 \cdot n^2 + c_1 \cdot n}$
R_2	$a_2 = 2 \cdot 1.4877 \cdot 10^{-7}$ $b_2 = 3 \cdot 2.9815 \cdot 10^{-3}$ $c_2 = 2.1 \cdot 15.0912$
	$\frac{3600 \cdot n}{a_2 \cdot n^3 - b_2 \cdot n^2 + c_2 \cdot n}$

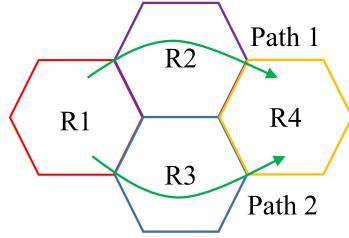


Fig. 6. A simple network.

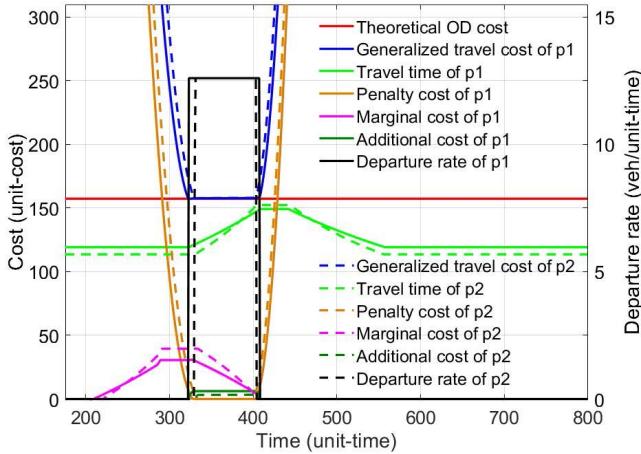


Fig. 5. Path departure rates against travel costs for the parallel network.

is $[0, 800]$ (unit-times) with regional inflow rate restricted to be less than or equal to 12.6 (unit/time). Expected arrival time interval is specified as $[t_{de}, t_{dl}] = [450, 550]$ (times). A quadratic early/late arrival penalty of the form (41) is imposed.

Path departure rate profiles are plotted against the generalized travel cost profiles and marginal cost profiles in Fig. 5. Neither of the paths is utilized at the very beginning since the cost exceeds the equilibrium travel cost. Both paths then become active when the equilibrium travel cost is reached. As these two paths are connecting the same OD pair, their equilibrium costs are the same and are exactly the same as the theoretical value as shown in Fig. 5. Vehicles traveling on Path 1 depart earlier than those on Path 2 regardless that the free-flow time of Path 2 is smaller than that of Path 1. This is because the dynamic marginal cost of Path 2 is significantly higher. The dynamic marginal cost, the time integral of the sensitivity value of the path cost with respect to a perturbation in the regional traffic accumulation during a vehicle's presence on the network, would appear before the departure flow.⁶ This is an important characteristic of the MFD model with time delay in flow propagation and the trip-based MFD model.

C. Two Simple Network Cases

Consider a simple network consisting of 4 regions with a single OD pair tied by 2 paths as shown in Fig. 6. The OD pair

⁶The first nontrivial marginal cost would appear at a free-flow time before the first departure flow.

TABLE III
NETWORK CONFIGURATION

Network configuration	Nonlinear network delay
R_1	$a_1 = 2 \cdot 1.4877 \cdot 10^{-7}$ $b_1 = 2 \cdot 2.9815 \cdot 10^{-3}$ $c_1 = 2 \cdot 15.0912$
R_2	$\frac{3600 \cdot n}{a_1 \cdot n^3 - b_1 \cdot n^2 + c_1 \cdot n}$
R_3	$a_3 = 2 \cdot 1.4877 \cdot 10^{-7}$ $b_3 = 2 \cdot 2.9815 \cdot 10^{-3}$ $c_3 = 2 \cdot 15.0912$
R_4	$\frac{3600 \cdot n}{a_3 \cdot n^3 - b_3 \cdot n^2 + c_3 \cdot n}$
	$a_4 = 2 \cdot 1.4877 \cdot 10^{-7}$ $b_4 = 2 \cdot 2.9815 \cdot 10^{-3}$ $c_4 = 2 \cdot 15.0912$
	$\frac{3600 \cdot n}{a_4 \cdot n^3 - b_4 \cdot n^2 + c_4 \cdot n}$

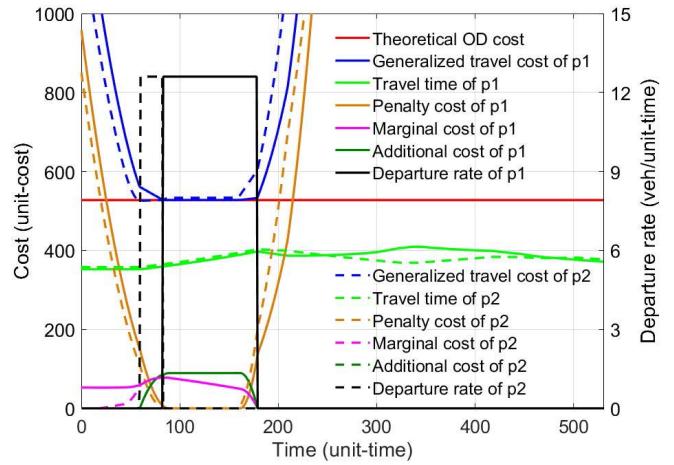


Fig. 7. Path departure rates against travel costs for the four region network.

from R_1 to R_4 is connected by 2 paths, $R_1 \rightarrow R_2 \rightarrow R_4$ (Path 1) and $R_1 \rightarrow R_3 \rightarrow R_4$ (Path 2) while the configuration of each region in conjunction with nonlinear network delay function is specified in Table III. The total travel amount between R_1 and R_4 is $Q_{od} = 1500$ (units). The planning horizon is $T = 800$ (unit-times) with regional inflow rate restricted to be less than or equal to 12.6 (unit/time). Expected arrival time interval is specified as $[t_{de}, t_{dl}] = [500, 700]$ (times). The early-late penalty function is the same as (41).

Solving the DSO problem for this simple network, path departure rate profiles are plotted against the corresponding travel cost profiles in Fig. 7. As indicated in the figure,

TABLE IV
NETWORK OD

OD	Path	Demand
OD1 ($R_1 \rightarrow R_7$)	$p_1 : R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_7$	3000
	$p_2 : R_1 \rightarrow R_4 \rightarrow R_7$	
	$p_3 : R_1 \rightarrow R_5 \rightarrow R_6 \rightarrow R_7$	

the equilibrium travel costs of the two paths are the same and is given by the theoretical OD travel cost. However, the departure time windows are significantly different for the two paths, even without any overlapping. As we can observe that, for the above three simulation cases, the generalized travel cost is very close to the theoretical OD travel cost. In the DTA literature, oscillation of the generalized travel cost can be observed even within the departure window regardless the path travel time is evaluated by the instantaneous journey time or experienced journey time, see e.g., [3], [5]. Such difference may be due to that we calculate the dynamic marginal cost in line with the interpretation of the marginal cost in the static SO analysis that the marginal cost is induced by a perturbation in the regional traffic accumulation (i.e., the link traffic volume in the static SO analysis). In [3], [5], it was postulated that the dynamic marginal cost is induced by a perturbation in the path departure flow profile. As the path departure flow profiles affect the regional traffic accumulations directly, the dynamic marginal cost would double-count the effect of a perturbation. As a matter of fact, there are two similar terms in the dynamic marginal cost when it is evaluated as a perturbation in the path departure flow profile. Of course, the effect of time discretization would also contribute to the oscillation in the DSO equilibrium results reported in the literature. The focus of this paper is on the properties of the DSO equilibrium under the MFD with time delay framework rather than on the numerical treatments. Discussing the effect of time discretization and efficient solution algorithms is out of the scope of this paper.

Next, for demonstration purpose, we solve the DSO for our motivated example as shown in Fig. 2(a), i.e., a larger network consisting of seven regions with a single OD pair. The OD pair connect R_1 to R_7 by three paths p_1, p_2, p_3 as summarized in Table IV. Configurations of the seven regions in conjunction with their network delay function are specified in Table V. The total travel demand between R_1 and R_7 is 3000 (units). The planning horizon is $T = 800$ (unit-times). Inflow rate to each region is restricted to be less than or equal to 12.6 (unit/time). Expected arrival time interval is specified as $[t_{de}, t_{dl}] = [400, 600]$ (times). The early-late penalty function is the same as (41).

We then solve the DSO problem for this network using the numerical algorithm depicted in the online appendix. Path departure rate profiles are plotted against the corresponding travel cost profiles in Fig. 8. The results indicate that the DSO equilibrium condition is well achieved. Only when the generalized travel cost equals the equilibrium cost, there will be departure for all the three paths. The results of these two simple network examples confirm the theoretical development of Proposition 3.1.

TABLE V
NETWORK CONFIGURATION

Region	Network configuration	Nonlinear network delay
R_1	$a_1 = 4 \cdot 1.4877 \cdot 10^{-7}$	$\frac{3600 \cdot n}{a_1 \cdot n^3 - b_1 \cdot n^2 + c_1 \cdot n}$
	$b_1 = 4 \cdot 2.9815 \cdot 10^{-3}$	
	$c_1 = 4 \cdot 15.0912$	
R_4	$a_2 = 2 \cdot 1.4877 \cdot 10^{-7}$	$\frac{3600 \cdot n}{a_2 \cdot n^3 - b_2 \cdot n^2 + c_2 \cdot n}$
	$b_2 = 2 \cdot 2.9815 \cdot 10^{-3}$	
	$c_2 = 2 \cdot 15.0912$	
R_5	$a_3 = 2 \cdot 1.4877 \cdot 10^{-7}$	$\frac{3600 \cdot n}{a_3 \cdot n^3 - b_3 \cdot n^2 + c_3 \cdot n}$
	$b_3 = 3.1 \cdot 2.9815 \cdot 10^{-3}$	
	$c_3 = 2.1 \cdot 15.0912$	

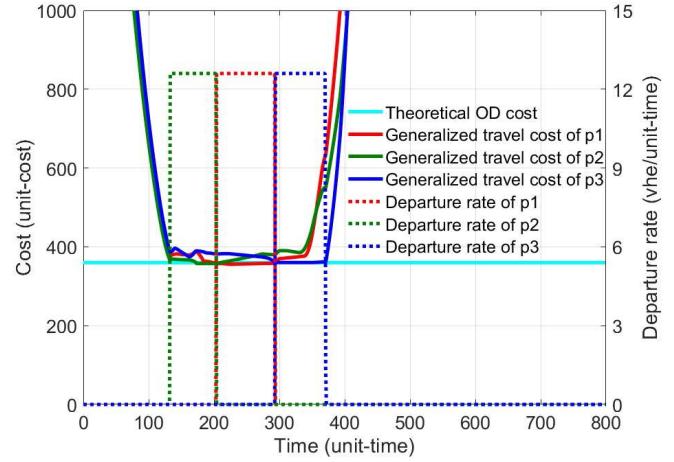


Fig. 8. Path departure rates against generalized travel cost for the seven-region network.

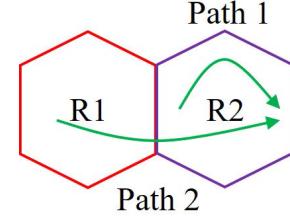


Fig. 9. An illustration of the network consists of two OD pairs.

D. A Simple Network With Two OD Pairs

We consider a simple network consisting of two regions with two OD pairs as shown in Fig. 9 to simulate the movements of the within-city commuters and suburban commuters in the morning peak. Configurations of the two regions in conjunction with network delay functions are specified in Table VII. The first OD pair is within region R_2 (i.e., the city center). The second OD pair is from the suburban region R_1 to the city center region R_2 . The OD travel demands and corresponding paths are summarized in Table VI. The destination of both paths is region R_2 . Both OD pairs admit the same expected arrival time window (EATW). The planning

TABLE VI
NETWORK ODS

OD	Path	Demand	EATW
OD1($R_2 \rightarrow R_2$)	$p_1 : R_2 \rightarrow R_2$	1000	[450, 550]
OD2($R_1 \rightarrow R_2$)	$p_2 : R_1 \rightarrow R_2$	500	[450, 550]

TABLE VII
NETWORK CONFIGURATION

	Network configuration	Nonlinear network delay
R_1	$a_1 = 2 \cdot 1.4877 \cdot 10^{-7}$	
	$b_1 = 2 \cdot 2.9815 \cdot 10^{-3}$	$\frac{3600 \cdot n}{a_1 \cdot n^3 - b_1 \cdot n^2 + c_1 \cdot n}$
	$c_1 = 2 \cdot 15.0912$	
R_2	$a_2 = 2 \cdot 1.4877 \cdot 10^{-7}$	
	$b_2 = 3 \cdot 2.9815 \cdot 10^{-3}$	$\frac{3600 \cdot n}{a_2 \cdot n^3 - b_2 \cdot n^2 + c_2 \cdot n}$
	$c_2 = 2.1 \cdot 15.0912$	

horizon is $T = 800$ (unit-times). The early-late penalty functions of the two OD pairs are assumed to be consistent with (41).

In the first case, we restrict the regional inflow rates to be less than or equal to 12.6 (unit/time) for both regions. The path departure rate of each OD pair and its corresponding travel cost are plotted in Fig. 10. The equilibrium costs differ between two OD pairs, i.e., different OD pairs would admit different minimum travel costs. For each OD pair, no traveler would depart if his/her generalized travel cost exceeds the corresponding equilibrium cost. This indicates that the DSO equilibrium condition is well-satisfied for the network with multiple OD pairs. Travelers traversing path 2 (from suburban region to the city center) depart earlier than path 1 despite the fact that the expected arrival time windows of both OD pairs are the same. This is because the travel time of path 2 is significant larger than that of path 1. Commuters who live closer to their place of employment have more degree of freedom to depart late. While for those long distance commuters (i.e., path 2), they have to depart early to catch the time and endures a higher (equilibrium) travel cost. This is consistent with the real world situations.

In this case, the path departure rate profiles do not overlap. Fig. 10 indicates that travelers of path 2 depart earlier to avoid the additional cost caused by the traffic restraint policy. To see the coordination between these two paths, we depict the total inflow rate to the destination region R_2 against the two components contributed by these paths in Fig. 11. Although travelers traversing path 2 depart from their origin much earlier, they enter region R_2 a bit later than the departure curve of path 1. The within-city commuters cooperate with the suburban commuters such that the travelers of both OD pairs would only suffer from the additional cost caused by the traffic restraint policy once either when they traverse region R_1 or region R_2 . However, this is subject to the expected arrival time window, the early-late penalty function, and the traffic restraint policy. To test the effect of the traffic restraint policy, the inflow rate to region R_1 is restricted to be less

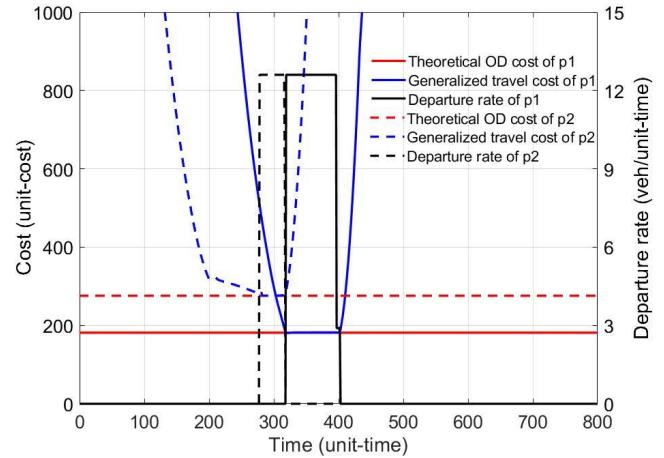


Fig. 10. Path departure rates and travel costs for the network with two OD pairs.

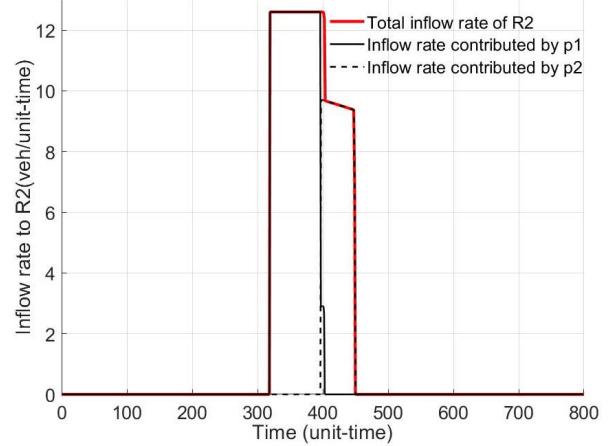


Fig. 11. Inflow rate to region R_2 and the corresponding contributions from the two paths.

than or equal to 6.3 (unit/time) while the traffic restraint plan of region R_2 remains unchanged. The path departure rate of each OD pair and the corresponding travel cost are plotted in Fig. 12. Due to a more restrictive traffic control in region R_1 , a larger departure time window is required for path 2 to serve the travel demand. As a result, the departure windows of the two OD pairs overlap. Some of the travelers on path 2 would suffer from the additional costs caused by the traffic restraint plans of both regions (see Fig. 13).

As indicated in (17), all paths would contribute to the marginal travel cost of a specific region if a path traverses the region regardless the OD pair it connects. Evaluating the marginal cost for a network with multiple OD pairs is more difficult than the single-OD case. This is also noticed in several recent papers, e.g., [4].

E. Comparison of the DSO and the DUE Under Different Expected Arrival Time Intervals

In this test, we consider the simulation scenario of Subsection IV-A. To test the effect of expected arrival time

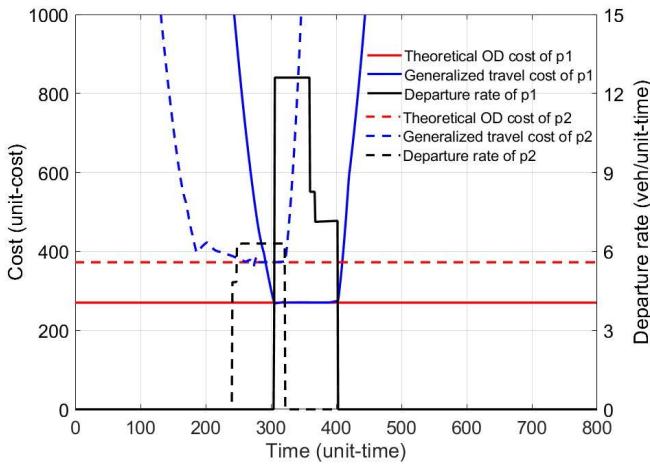


Fig. 12. Path departure rates and travel costs for the network with two OD pairs.

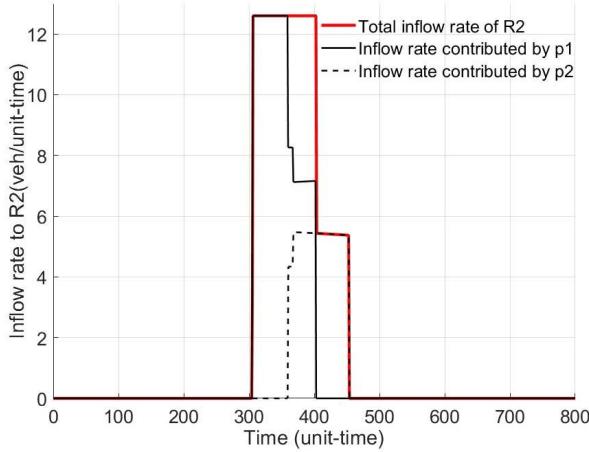


Fig. 13. Inflow rate to region R_2 and the corresponding contributions from the two paths.

interval, we assume that the expected arrival time is 500 unit-time, and vehicles are allowed to arrive early or late in a time window TW including this time instant, i.e the expected arrival time interval is $[t_{de}, t_{dl}] = [500 - TW/2, 500 + TW/2]$ (unit-time) with an early/late arrival penalty function (41).

The total travel cost of the DSO case is compared with that of the DUE case⁷ under different expected arrival time intervals. As shown in Table VIII, the total travel costs of both the DSO and the DUE decrease with respect to the increase in the width of the expected arrival time interval. This is because that when the expected arrival time interval is larger, travelers could depart in a more scattered manner and thus suffer less congestion externality. The gap between the DSO and the DUE also decreases with respect to the increase in the width of the expected arrival time interval as illustrated in Fig. 14. When the expected arrival time interval is large enough, the DSO is close to the DUE. This is because travelers can depart sparsely in time when the expected arrival time interval is large enough

⁷The DUE case was solved using the formulation and solution algorithm proposed in [39], [53].

TABLE VIII
COMPARISON OF DSO AND DUE COSTS UNDER DIFFERENT EXPECTED ARRIVAL TIME INTERVALS

Time window (unit-time)	Total travelers' cost		Gap
	DSO	DUE	
100	$1.32 \cdot 10^6$	$1.53 \cdot 10^6$	16.45%
120	$1.05 \cdot 10^6$	$1.22 \cdot 10^6$	16.11%
140	$8.23 \cdot 10^5$	$9.51 \cdot 10^5$	15.44%
160	$6.30 \cdot 10^5$	$7.26 \cdot 10^5$	15.24%
180	$4.72 \cdot 10^5$	$5.35 \cdot 10^5$	13.42%
200	$3.49 \cdot 10^5$	$3.91 \cdot 10^5$	12.24%

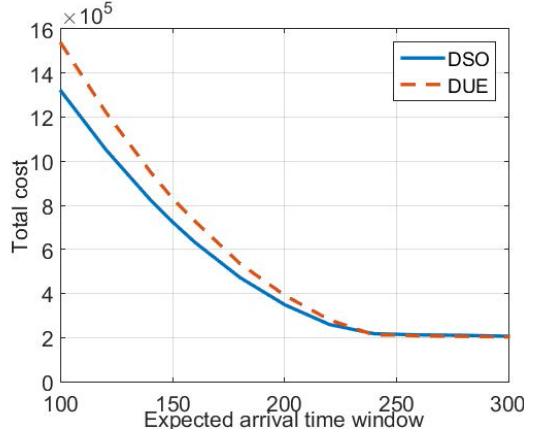


Fig. 14. The DSO against the DUE with respect to different expected arrival time intervals.

such that no congestion would occur. Both the DSO and the DUE equilibrium costs (in terms of travel time) are close to the free-flow travel time.

V. CONCLUSIONS

In this paper, the dynamic system optimum (DSO) problem with simultaneous route and departure time assignments for a general network modeled by multi-region MFD systems was formulated. The DSO problem was captured by a constrained optimal control framework. To overcome the limitation of inconsistent flow propagation between region boundaries of the conventional MFD model, this paper explicitly considers the regional travel time in the MFD model based on the flow conservation law. Dynamic equilibrium condition was derived analytically under the umbrella of the Pontryagin minimum principle. The structure of path-specific marginal cost was analyzed regarding to the early-late penalty function. Inflow capacity constraints are imposed to prevent the network from over-saturation and thus the “gridlock”. The additional cost caused by the inflow capacity constraint can be regarded as the price to prevent the network entering over-saturated conditions. In contrast to existing analytical methods, the proposed method is applicable for general MFD systems without local linearization of the MFD dynamics nor approximation of the equilibrium solution. Numerical examples were conducted to illustrate the characteristics of DSO traffic equilibrium and corresponding marginal cost together with other dynamic external costs. Implications for the network-level dynamic road

pricing design were also discussed in the numerical studies. In the future work, we will extend the proposed framework to analyze the dynamic system optimal problem for traffic mixed with connected automated vehicles and regular human-piloted vehicles. Efficient numerical schemes for calculating the dynamic marginal travel cost for general networks are also challenging future works.

REFERENCES

- [1] M. Kuwahara, "A theory and implications on dynamic marginal cost," *Transp. Res. A, Policy Pract.*, vol. 41, no. 7, pp. 627–643, Aug. 2007.
- [2] A. H. F. Chow, "Dynamic system optimal traffic assignment—A state-dependent control theoretic approach," *Transportmetrica*, vol. 5, no. 2, pp. 85–106, May 2009.
- [3] R. Zhong, A. Sumalee, and T. Maruyama, "Dynamic marginal cost, access control, and pollution charge: A comparison of bottleneck and whole link models," *J. Adv. Transp.*, vol. 46, no. 3, pp. 191–221, Jul. 2012.
- [4] J. Long and W. Y. Szeto, "Link-based system optimum dynamic traffic assignment problems in general networks," *Oper. Res.*, vol. 67, no. 1, pp. 167–182, Jan. 2019.
- [5] A. H. F. Chow, "Properties of system optimal traffic assignment with departure time choice and its solution method," *Transp. Res. B, Methodol.*, vol. 43, no. 3, pp. 325–344, Mar. 2009.
- [6] W. Shen and H. M. Zhang, "On the morning commute problem in a corridor network with multiple bottlenecks: Its system-optimal traffic flow patterns and the realizing tolling scheme," *Transp. Res. B, Methodol.*, vol. 43, no. 3, pp. 267–284, Mar. 2009.
- [7] H. M. Zhang and W. Shen, "Access control policies without inside queues: Their properties and public policy implications," *Transp. Res. B, Methodol.*, vol. 44, nos. 8–9, pp. 1132–1147, Sep. 2010.
- [8] M. O. Ghali and M. J. Smith, "A model for the dynamic system optimum traffic assignment problem," *Transp. Res. B, Methodol.*, vol. 29, no. 3, pp. 155–170, Jun. 1995.
- [9] R. Arnott and M. Kraus, "When are anonymous congestion charges consistent with marginal cost pricing?" *J. Public Econ.*, vol. 67, no. 1, pp. 45–64, Jan. 1998.
- [10] Y. Liu, Y. J. Nie, and J. Hall, "A semi-analytical approach for solving the bottleneck model with general user heterogeneity," *Transp. Res. B, Methodol.*, vol. 71, pp. 56–70, Jan. 2015.
- [11] D. K. Merchant and G. L. Nemhauser, "A model and an algorithm for the dynamic traffic assignment problems," *Transp. Sci.*, vol. 12, no. 3, pp. 183–199, Aug. 1978.
- [12] D. K. Merchant and G. L. Nemhauser, "Optimality conditions for a dynamic traffic assignment model," *Transp. Sci.*, vol. 12, no. 3, pp. 200–207, Aug. 1978.
- [13] A. K. Ziliaskopoulos, "A linear programming model for the single destination system optimum dynamic traffic assignment problem," *Transp. Sci.*, vol. 34, no. 1, pp. 37–49, Feb. 2000.
- [14] Y. M. Nie, "A cell-based Merchant–Nemhauser model for the system optimum dynamic traffic assignment problem," *Transp. Res. B, Methodol.*, vol. 45, no. 2, pp. 329–342, Feb. 2011.
- [15] F. Zhu and S. V. Ukkusuri, "A cell based dynamic system optimum model with non-holding back flows," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 367–380, Nov. 2013.
- [16] H. Zheng, Y.-C. Chiu, and P. B. Mirchandani, "On the system optimum dynamic traffic assignment and earliest arrival flow problems," *Transp. Sci.*, vol. 49, no. 1, pp. 13–27, Feb. 2015.
- [17] Z. S. Qian, W. Shen, and H. Zhang, "System-optimal dynamic traffic assignment with and without queue spillback: Its path-based formulation and solution via approximate path marginal cost," *Transp. Res. B*, vol. 46, no. 7, pp. 874–893, 2012.
- [18] R. Ma, X. J. Ban, and J.-S. Pang, "Continuous-time dynamic system optimum for single-destination traffic networks with queue spillbacks," *Transp. Res. B, Methodol.*, vol. 68, pp. 98–122, Oct. 2014.
- [19] R. Ma, X. J. Ban, and W. Y. Szeto, "Emission modeling and pricing on single-destination dynamic traffic networks," *Transp. Res. B, Methodol.*, vol. 100, pp. 255–283, Jun. 2017.
- [20] K. Han, T. L. Friesz, and T. Yao, "A partial differential equation formulation of Vickrey's bottleneck model, part I: Methodology and theoretical analysis," *Transp. Res. B, Methodol.*, vol. 49, pp. 55–74, Mar. 2013.
- [21] T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B. W. Wie, "A variational inequality formulation of the dynamic network user equilibrium problem," *Oper. Res.*, vol. 41, no. 1, pp. 179–191, Feb. 1993.
- [22] K. Han, W. Y. Szeto, and T. L. Friesz, "Formulation, existence, and computation of boundedly rational dynamic user equilibrium with fixed or endogenous user tolerance," *Transp. Res. B, Methodol.*, vol. 79, pp. 16–49, Sep. 2015.
- [23] K. Han, T. L. Friesz, W. Y. Szeto, and H. Liu, "Elastic demand dynamic network user equilibrium: Formulation, existence and computation," *Transp. Res. B, Methodol.*, vol. 81, pp. 183–209, Nov. 2015.
- [24] K. Han, B. Piccoli, and W. Y. Szeto, "Continuous-time link-based kinematic wave model: Formulation, solution existence, and well-posedness," *Transportmetrica B, Transp. Dyn.*, vol. 4, no. 3, pp. 187–222, Sep. 2016.
- [25] M. Amirgholy and H. O. Gao, "Modeling the dynamics of congestion in large urban networks using the macroscopic fundamental diagram: User equilibrium, system optimum, and pricing strategies," *Transp. Res. B, Methodol.*, vol. 104, pp. 215–237, Oct. 2017.
- [26] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transp. Res. B, Methodol.*, vol. 41, no. 1, pp. 49–62, Jan. 2007.
- [27] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transp. Res. B, Methodol.*, vol. 42, no. 9, pp. 771–781, Nov. 2008.
- [28] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transp. Res. B, Methodol.*, vol. 42, no. 9, pp. 759–770, Nov. 2008.
- [29] V. L. Knoop, S. P. Hoogendoorn, and J. W. C. Van Lint, "Routing strategies based on macroscopic fundamental diagram," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2315, no. 1, pp. 1–10, Jan. 2012.
- [30] L. Leclercq and N. Geroliminis, "Estimating MFDs in simple networks with route choice," *Transp. Res. B, Methodol.*, vol. 57, pp. 468–484, Nov. 2013.
- [31] C. F. Daganzo, V. V. Gayah, and E. J. Gonzales, "Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability," *Transp. Res. Part B: Methodol.*, vol. 45, no. 1, pp. 278–288, Jan. 2011.
- [32] K. Wada, K. Satsukawa, M. Smith, and T. Akamatsu, "Network throughput under dynamic user equilibrium: Queue spillback, paradox and traffic control," *Transp. Res. B, Methodol.*, vol. 126, pp. 391–413, Aug. 2019.
- [33] R. X. Zhong, Y. P. Huang, C. Chen, W. H. K. Lam, D. B. Xu, and A. Sumalee, "Boundary conditions and behavior of the macroscopic fundamental diagram based network traffic dynamics: A control systems perspective," *Transp. Res. B, Methodol.*, vol. 111, pp. 327–355, May 2018.
- [34] M. Yildirimoglu and N. Geroliminis, "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams," *Transp. Res. B, Methodol.*, vol. 70, pp. 186–200, Dec. 2014.
- [35] M. Yildirimoglu, M. Ramezani, and N. Geroliminis, "Equilibrium analysis and route guidance in large-scale networks with MFD dynamics," *Transp. Res. Proc.*, vol. 9, pp. 185–204, 2015.
- [36] K. Ampountolas, N. Zheng, and N. Geroliminis, "Macroscopic modelling and robust control of bi-modal multi-region urban road networks," *Transp. Res. B, Methodol.*, vol. 104, pp. 616–637, Oct. 2017.
- [37] M. Yildirimoglu, I. I. Sirmatel, and N. Geroliminis, "Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance," *Transp. Res. B, Methodol.*, vol. 118, pp. 106–123, Dec. 2018.
- [38] R. Aghamohammadi and J. A. Laval, "Dynamic traffic assignment using the macroscopic fundamental diagram: A review of vehicular and pedestrian flow models," *Transp. Res. B, Methodol.*, Nov. 2018.
- [39] Y. P. Huang *et al.*, "A dynamic user equilibrium model for multi-region macroscopic fundamental diagram systems with time-varying delays," *Transp. Res. B, Methodol.*, vol. 131, pp. 1–25, Jan. 2020.
- [40] J. Haddad and B. Mirkin, "Adaptive perimeter traffic control of urban road networks based on MFD model with time delays," *Int. J. Robust Nonlinear Control*, vol. 26, no. 6, pp. 1267–1285, Apr. 2016.
- [41] J. Haddad and Z. Zheng, "Adaptive perimeter control for multi-region accumulation-based models with state delays," *Transp. Res. B, Methodol.*, vol. 26, pp. 1267–1285, Jul. 2018.
- [42] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transp. Res. B, Methodol.*, vol. 46, no. 10, pp. 1393–1403, Dec. 2012.
- [43] K. Aboudolas and N. Geroliminis, "Perimeter and boundary flow control in multi-reservoir heterogeneous networks," *Transp. Res. B, Methodol.*, vol. 55, pp. 265–281, Sep. 2013.

- [44] R. Arnott, "A bathtub model of downtown traffic congestion," *J. Urban Econ.*, vol. 76, pp. 110–121, Jul. 2013.
- [45] G. Mariotte, L. Leclercq, and J. A. Laval, "Macroscopic urban dynamics: Analytical and numerical comparisons of existing models," *Transp. Res. B, Methodol.*, vol. 101, pp. 245–267, Jul. 2017.
- [46] R. X. Zhong, C. Chen, Y. P. Huang, A. Sumalee, W. H. K. Lam, and D. B. Xu, "Robust perimeter control for two urban regions with macroscopic fundamental diagrams: A control-Lyapunov function approach," *Transp. Res. Proc.*, vol. 23, pp. 922–941, 2017.
- [47] R. Arnott and J. Buli, "Solving for equilibrium in the basic bathtub model," *Transp. Res. B, Methodol.*, vol. 109, pp. 150–175, Mar. 2018.
- [48] X. Ban, H. X. Liu, M. C. Ferris, and B. Ran, "A link-node complementarity model and solution algorithm for dynamic user equilibria with exact flow propagations," *Transp. Res. B, Methodol.*, vol. 42, no. 9, pp. 823–842, Nov. 2008.
- [49] T. L. Friesz, D. Bernstein, Z. Suo, and R. L. Tobin, "Dynamic network user equilibrium with state-dependent time lags," *Netw Spat Econ*, vol. 1, no. 3, pp. 319–347, 2001.
- [50] T. L. Friesz and K. Han, "The mathematical foundations of dynamic user equilibrium," *Transp. Res. B, Methodol.*, vol. 126, pp. 309–328, Aug. 2019.
- [51] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.
- [52] H. S. Mahmassani, M. Saberi, and A. Zockaei, "Urban network gridlock: Theory, characteristics, and dynamics," *Transport. Res. C*, vol. 36, pp. 480–497, Dec. 2013.
- [53] R. Zhong, Y. Huang, J. Xiong, N. Zheng, W. Lam, and A. Sumalee, "An optimal control framework for multi-region macroscopic fundamental diagram systems with time delay, considering route choice and departure time choice," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1962–1967.



Renxin Zhong is currently an Associate Professor with the School of Intelligent Systems Engineering, Sun Yat-sen University. He is also the Deputy Director of the Guangdong Key Laboratory of Intelligent Transportation Systems. His main research interests include traffic incident detection and management strategies, machine learning and data mining for transportation big data analysis, and optimal and nonlinear control theory with applications in transportation engineering. He received the Outstanding Dissertation Paper Award and the Gordon Newell Memorial Prize at the 17th HKSTS International Conference, as well as the First Runner-up of the HKSTS Outstanding Student Paper Award at the 14th HKSTS International Conference. His article was shortlisted for the Best Paper Award at the IEEE ITSC 2018. His team won third place in the KDD Cup 2017 (freeway travel time and traffic flow estimations) and KDD Cup 2018 (urban pollutant prediction). He serves as an Associate Editor for *Transportmetrica A* and a Guest Editor for *Transportmetrica B*.

Memorial Prize at the 17th HKSTS International Conference, as well as the First Runner-up of the HKSTS Outstanding Student Paper Award at the 14th HKSTS International Conference. His article was shortlisted for the Best Paper Award at the IEEE ITSC 2018. His team won third place in the KDD Cup 2017 (freeway travel time and traffic flow estimations) and KDD Cup 2018 (urban pollutant prediction). He serves as an Associate Editor for *Transportmetrica A* and a Guest Editor for *Transportmetrica B*.



Jianhui Xiong received the B.S. degree in traffic engineering from Sun Yat-sen University, Guangzhou, China, in 2017, where he is currently pursuing the M.Phil. degree in transportation engineering. His research interests include transportation systems modeling and control, and data mining for intelligent transportation systems.



Yunping Huang received the B.S. degree in traffic engineering and the master's degree in transportation planning and management from Sun Yat-sen University, Guangzhou, China, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree in transportation engineering with The Hong Kong Polytechnic University. She has published several articles in *Transportation Research Part B* and ISTTT. Her research interests include transportation systems modeling and control, and public transit systems.



Agachai Sumalee was a Full Professor in transportation engineering with The Hong Kong Polytechnic University. He is currently a Chair Professor with the School of Integrated Innovation, Chulalongkorn University. His research areas are transit planning, intelligent transport system, network modeling, and optimization and transport economics. He has published more than 100 research articles in top peer-reviewed journals on these research topics. He has received several prizes and awards for his research works, including the APEC Science Prize for Innovation, Research and Education, and the National Innovation Awards of Thailand. He is also the Editor-in-Chief of *Transportmetrica B*, an Associate Editor of *Networks and Spatial Economics*, and an editorial board member of several prestigious journals.



Andy H. F. Chow received the Ph.D. degree in transport studies from University College London (UCL), London, U.K., in 2007. He was a Senior Lecturer in transport studies with UCL, a Post-Doctoral Researcher with the University of California Berkeley, under the California PATH Program. He is currently an Assistant Professor with the Department of Architecture and Civil Engineering, City University of Hong Kong (CityU). His research focuses on modeling and optimizing dynamic transport systems. He received the Gordon Newell Memorial Dissertation Prize in 2008 for his doctoral dissertation on optimal control of dynamic transport networks.



Tianlu Pan received the B.Eng. degree (Hons.) in electronic engineering from Sun Yat-sen University in 2007, and the M.Phil. and Ph.D. degrees in transportation engineering from The Hong Kong Polytechnic University in 2011 and 2018, respectively.

She is currently with the Guangdong Key Laboratory of Urban Informatics and the Shenzhen Key Laboratory of Spatial Smart Sensing and the Services, School of Architecture and Urban Planning, Research Institute for Smart Cities, Shenzhen University. Her research interests include dynamic traffic surveillance and management for intelligent transportation systems applications, network modeling and optimization especially for the emerging vehicle automation and communication systems, and network reliability and performance assessment.